DOCUMENT RESUME

ED 410 539 CS 012 900

AUTHOR	Campbell, Jay R.; Donahue, Patricia L.
TITLE	Students Selecting Stories: The Effects of Choice in Reading Assessment. Results from "The NAEP Reader" Special Study of the 1994 National Assessment of Educational Progress.
INSTITUTION	National Assessment of Educational Progress, Princeton, NJ.; Educational Testing Service, Princeton, NJ.
SPONS AGENCY	National Center for Education Statistics (ED), Washington, DC.
REPORT NO	NCES-97-491
PUB DATE	1997-06-00
NOTE	106p.; The title page title says "Special Study"; the cover title says "special Survey."
AVAILABLE FROM	National Library of Education, Office of Educational Research and improvement, U.S. Department of Education, 555 New Jersey Avenue, N.W., Washington, DC 20208-5641; phone: 1-800-424-1616; World Wide Web: http://www.ed.gov/NCES/naep
PUB TYPE	Reports - Research (143)
EDRS PRICE	MF01/PC05 Plus Postage.
DESCRIPTORS	Comparative Analysis; Grade 12; Grade 8; *Reading Achievement; *Reading Material Selection; Reading Research; Secondary Education; Sex Differences; *Story Reading; *Student Attitudes; *Student Evaluation
IDENTIFIERS	*National Assessment of Educational Progress

ABSTRACT

Conducted as part of the 1994 National Assessment of Educational Progress (NAEP), the "NAEP Reader" study was designed to compare the performance of students who were allowed to select a story with the performance of those who were assigned a story. One nationally representative sample of students at each grade was allowed to choose a story (from 7) to read. Distinct representative samples at each grade were assigned stories. All participants answered the same 11 comprehension questions. Results indicated: (1) no significant difference was observed among 12th- graders (and only a slight difference among 8th-graders) between the average reading scores of students who were given a choice of story and students who were assigned a story; (2) at grades 8 and 12, students in the choice group were more likely than students in the non-choice group to rate the assessment as easier than other tests or assignments that they had in school; (3) despite slight variations, the patterns of students' story selection were similar across racial/ethnic groups at both grades 8 and 12, and across gender groups at grade 8; (4) among 12th-graders, males were drawn to a story about a soldier and females were drawn to a story about a relationship; and (5) although identical questions were used, many questions were more or less difficult to answer in conjunction with certain stories. (Contains 14 tables and 10 figures of data. Appendixes contain data; the comprehension questions; and details of the procedures and methods used.)(RS)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 410 539

NATIONAL CENTER FOR EDUCATION STATISTICS

Students Selecting Stories:
The Effects of Choice in Reading Assessment

Results from The *NAEP Reader* Special Survey
of the 1994 National Assessment of Educational Progress

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

CS 012 900

BEST COPY AVAILABLE

THE NATION'S REPORT CARD naep

Prepared by Educational Testing Service under a cooperative agreement with the National Center for Education Statistics

U.S. Department of Education
Office of Educational Research and Improvement NCES 97-491

What is The Nation's Report Card?

THE NATION'S REPORT CARD, the National Assessment of Educational Progress (NAEP), is the only nationally representative and continuing assessment of what America's students know and can do in various subject areas. Since 1969, assessments have been conducted periodically in reading, mathematics, science, writing, history/geography, and other fields. By making objective information on student performance available to policymakers at the national, state, and local levels, NAEP is an integral part of our nation's evaluation of the condition and progress of education. Only information related to academic achievement is collected under this program. NAEP guarantees the privacy of individual students and their families.

NAEP is a congressionally mandated project of the National Center for Education Statistics, the U.S. Department of Education. The Commissioner of Education Statistics is responsible, by law, for carrying out the NAEP project through competitive awards to qualified organizations. NAEP reports directly to the Commissioner, who is also responsible for providing continuing reviews, including validation studies and solicitation of public comment, on NAEP's conduct and usefulness.

In 1988, Congress established the National Assessment Governing Board (NAGB) to formulate policy guidelines for NAEP. The Board is responsible for selecting the subject areas to be assessed from among those included in the National Education Goals; for setting appropriate student performance levels; for developing assessment objectives and test specifications through a national consensus approach; for designing the assessment methodology; for developing guidelines for reporting and disseminating NAEP results; for developing standards and procedures for interstate, regional, and national comparisons; for determining the appropriateness of test items and ensuring they are free from bias; and for taking actions to improve the form and use of the National Assessment.

The National Assessment Governing Board

Honorable William T. Randall, Chair
Former Commissioner of Education
State of Colorado
Denver, Colorado

Mary R. Blanton, Vice Chair
Attorney
Salisbury, North Carolina

Patsy Cavazos
Principal
W.G. Love Accelerated Elementary School
Houston, Texas

Catherine A. Davidson
Secondary Education Director
Central Kitsap School District
Silverdale, Washington

Edward Donley
Former Chairman
Air Products & Chemicals, Inc.
Allentown, Pennsylvania

Honorable James Edgar
Member Designate
Governor of Illinois
Springfield, Illinois

James E. Ellingson
Fourth-Grade Classroom Teacher
Probstfield Elementary School
Moorhead, Minnesota

Thomas H. Fisher
Director, Student Assessment Services
Florida Department of Education
Tallahassee, Florida

Michael J. Guerra
Executive Director
Secondary Schools Department
National Catholic Educational Association
Washington, DC

Edward H. Haertel
Professor of Education
Stanford University
Stanford, California

Jan B. Loveless
District Communications Specialist
Midland Public Schools
Midland, Michigan

Marilyn McConachie
Former School Board Member
Glenbrook High Schools
Glenview, Illinois

William J. Moloney
Superintendent of Schools
Calvert County Public Schools
Prince Frederick, Maryland

Honorable Annette Morgan
Former Member
Missouri House of Representatives
Jefferson City, Missouri

Mark D. Musick
President
Southern Regional Education Board
Atlanta, Georgia

Mitsugi Nakashima
First Vice-Chairperson
Hawaii State Board of Education
Honolulu, Hawaii

Michael T. Nettles
Professor of Education & Public Policy
University of Michigan
Ann Arbor, Michigan
and Director
Frederick D. Patterson Research Institute
United Negro College Fund

Honorable Norma Paulus
Superintendent of Public Instruction
Oregon State Department of Education
Salem, Oregon

Honorable Roy Romer
Governor of Colorado
Denver, Colorado

Honorable Edgar D. Ross
Judge
Territorial Court of the Virgin Islands
Christiansted, St. Croix
U.S. Virgin Islands

Fannie L. Simmons
Mathematics Coordinator
District 5 of Lexington/Richland County
Ballentine, South Carolina

Adam Urbanski
President
Rochester Teachers Association
Rochester, New York

Deborah Voltz
Assistant Professor
Department of Special Education
University of Louisville
Louisville, Kentucky

Marilyn A. Whirry
Twelfth-Grade English Teacher
Mira Costa High School
Manhattan Beach, California

Dennie Palmer Wolf
Senior Research Associate
Harvard Graduate School of Education
Cambridge, Massachusetts

Ramon C. Cortines (Ex-Officio)
Acting Assistant Secretary
Office of Educational Research
and Improvement
U.S. Department of Education
Washington, DC

Roy Truby
Executive Director, NAGB
Washington, DC

NATIONAL CENTER FOR EDUCATION STATISTICS

Students Selecting Stories: The Effects of Choice in Reading Assessment

Results from *The NAEP Reader*
Special Study of the
1994 National Assessment of Educational Progress

Jay R. Campbell
Patricia L. Donahue

June 1997

U.S. Department of Education
Office of Educational Research and Improvement **NCES 97-491**

Prepared by Educational Testing Service under a cooperative agreement with the National Center for Education Statistics

U.S. Department of Education
Richard W. Riley
Secretary

Office of Educational Research and Improvement
Ramon C. Cortines
Acting Assistant Secretary

National Center for Education Statistics
Pascal D. Forgione, Jr.
Commissioner

Education Assessment Group
Gary W. Phillips
Associate Commissioner

June 1997

SUGGESTED CITATION
Campbell, J.R., & Donahue, P.L. (1997).
Students selecting stores: The effects of choice in reading assessment.
Washington, DC: National Center for Education Statistics.

FOR MORE INFORMATION
Contact:
Arnold A. Goldstein
202-219-1741

For ordering information on this report, write:

National Library of Education
Office of Educational Research and Improvement
U.S. Department of Education
555 New Jersey Avenue, NW
Washington, D.C. 20208-5641

or call 1-800-424-1616 (in the Washington, DC, metropolitan area call 202-219-1651).

This report also is available on the World Wide Web: http://www.ed.gov/NCES/naep

The work upon which this publication is based was performed for the National Center for Education Statistics, Office of Educational Research and Improvement, by Educational Testing Service.

Educational Testing Service is an equal opportunity, affirmative action employer.

Educational Testing Service, ETS, and *the ETS logo* are registered trademarks of Educational Testing Service.

Contents

Introduction 1

Major Findings 2
Background on the NAEP Reading Assessment 3
Rationale for *The NAEP Reader* Study 4
Design of *The NAEP Reader* Study 6
Figure 1 Eighth-Grade NAEP Reader Story Summaries 8
Figure 2 Twelfth-Grade NAEP Reader Story Summaries 9
This Report 11

Chapter 1: Choice vs. Non-Choice: Student Performance and Perceptions 13

Table 1.1 Choice vs. Non-Choice Average Scale Scores for Eighth-Grade Students 14
Table 1.2 Choice vs. Non-Choice Average Scale Scores for Twelfth-Grade Students 15
Table 1.3 Average Scale Scores of Eighth- and Twelfth-Grade Students in the Choice and Non-Choice Samples by Story 16
Table 1.4 Percentages of Students Not Reaching the Last Two Comprehension Questions 17
Students' Perceptions of the Assessment and Their Motivation 18
Table 1.5 Students' Perceptions of How Many Questions They Got Right 18
Table 1.6 Students' Perceptions of How Hard This Test Was Compared to Other Reading Tests or Assignments 19
Table 1.7 Students' Reports About How Hard They Tried Compared to Other Reading Tests or Assignments 20
Table 1.8 Students' Reports of How Important It Was to Them to Do Well 21
Summary and Discussion 22

Chapter 2: Patterns of Students' Story Selections 23

The NAEP Reader: Grade 8 24
The NAEP Reader: Grade 12 26
Overall Pattern of Students' Selections of Stories in *The NAEP Reader* 29

Table 2.1 Percentages of Students Selecting Each Story in *The NAEP Reader* 29
Pattern of Students' Story Selections by Race/Ethnicity 30
Table 2.2 Percentages of Eighth-Grade Students Selecting Each Story in *The NAEP Reader* by Race/Ethnicity 30
Table 2.3 Percentages of Twelfth-Grade Students Selecting Each Story in *The NAEP Reader* by Race/Ethnicity 31
Pattern of Students' Story Selections by Gender 32
Table 2.4 Percentages of Eighth-Grade Students Selecting Each Story in *The NAEP Reader* by Gender 32
Table 2.5 Percentages of Twelfth-Grade Students Selecting Each Story in *The NAEP Reader* by Gender 33
Students' Criteria for Selecting Stories 33
Figure 2.1 Coding Categories for Describing Students' Story Selection Criteria 34
Table 2.6 Students' Reports of Story Selection Criteria 35
Summary and Discussion 36

Chapter 3: Context Effects on Constructed-Response Questions: Variations in Students' Responses Across Stories 39
Eighth Graders' Performance on Comprehension Questions Across Stories 41
Figure 3.1 Percentages of Acceptable Responses to Short Constructed-Response Question #10 Across the Seven Stories at Grade 8 41
Figure 3.2 Percentages of Acceptable Responses to Short Constructed-Response Question #2 Across the Seven Stories at Grade 8 44
Figure 3.3 Percentages of Essential or Better Responses to Extended Constructed-Response Question #11 Across the Seven Stories at Grade 8 47
Twelfth Graders' Performance on Comprehension Questions Across Stories 50
Figure 3.4 Percentages of Acceptable Responses to Short Constructed Response Question #10 Across the Seven Stories at Grade 12 50
Figure 3.5 Percentages of Acceptable Responses to Short Constructed Response Question #7 Across the Seven Stories at Grade 12 53
Figure 3.6 Percentages of Essential or Better Responses to Extended Constructed-Response Question #11 Across the Seven Stories at Grade 12 56
Figure 3.7 Percentages of Acceptable Responses to Short Constructed Response Question #12 Across the Seven Stories at Grade 12 59
Summary and Discussion 60

Conclusion 61

Summary of Findings 61

Study Design Issues and Alternative Methods 63

Appendix A: The NAEP Reader Comprehension Questions 67

Appendix B: Procedures and Methods 73

Developing *The NAEP Reader* Study 74

Figure B.1 The Reading Instrument Development Committee 74

Design of the 1994 *NAEP Reader* Study 75

Sampling 76

Table B.1 Weighted Choice and Non-Choice Sample Sizes by Grade and Story 76

Table B.2 Sample Sizes and Weighted Percentages of Students by Race/Ethnicity and Gender in the Non-Choice Samples for Each Story and in the Total Non-Choice and Choice Samples: Grade 8 77

Table B.3 Sample Sizes and Weighted Percentages of Students by Race/Ethnicity and Gender in the Non-Choice Samples for Each Story and in the Total Non-Choice and Choice Samples: Grade 12 78

NAEP Reader Reporting Groups 79

Scoring 79

Table B.4 Percentages of Exact Agreement Between First and Second Scorers on the Reliability Sample of Responses to Comprehension Questions 81

Data Analysis and IRT Scoring 82

Estimating Variability 83

Drawing Inferences from the Results 84

Appendix C: Student Performance on Comprehension Questions: Grade 8 87

Appendix D: Student Performance on Comprehension Questions: Grade 12 93

Acknowledgments

Introduction

Students select stories to read in classrooms, in libraries, in their homes, and wherever reading materials are available. They choose different types of stories based on their purpose for reading and their personal interests. One situation where students rarely have a choice of what to read, however, is in a reading assessment. Some educators view this as a problem in assessing reading comprehension since students may be more engaged when they have chosen a text than when they are reading assigned texts.

While the effect of choice on student performance in large-scale assessments and the psychometric ramifications of offering choice have been studied, *The NAEP Reader* study represents a significant departure from past efforts. Whereas earlier studies have focused on the effects of offering students a choice among test questions,[1] *The NAEP Reader* study was designed to examine the feasibility and measurement impact of offering test takers a choice of reading material on an assessment of reading comprehension.

Conducted as part of the 1994 National Assessment of Educational Progress (NAEP), *The NAEP Reader* study was designed to compare the performance of students who were allowed to select a story with the performance of those who were assigned a story. Booklets containing a selection of seven stories were produced, one for grade 8 and a different selection for grade 12. One nationally representative sample of students at each grade was allowed to choose a story to read. Distinct representative samples at each grade were assigned stories. As *The NAEP Reader* study was administered in conjunction with the NAEP reading assessment, students participating in the study worked within the same 50-minute time frame as students taking the main assessment. All participants, in both the choice and non-choice samples, answered the same eleven comprehension questions that were generically worded so as to be applicable to each and all of the stories. Students in the choice sample were given an additional question asking them to briefly explain the reason for their choice of story. The major findings from this special study are provided below.

---

[1] Lukhele, R., Thissen, D. & Wainer, H. (1994). On the relative value of multiple-choice, constructed response, and examinee-selected items on two achievement tests. *Journal of Educational Measurement, 31*(3), 234-250.

Major Findings

◆ *Choice vs. Non-Choice Performance*

Among twelfth graders, no significant difference was observed between the average reading scores of students who were given a choice of story and students who were assigned a story. At grade 8, however, students who selected a story demonstrated slightly lower performance than students who did not have a choice of story. The difference was one scale-score point on a 0-to-100 scale with a standard deviation of 10.

◆ *Choice vs. Non-Choice Perceptions*

Some differences were observed in students' perceptions of the assessment depending on whether or not they were allowed to choose a story. At both grades 8 and 12, students in the choice group were more likely than students in the non-choice group to rate the assessment as easier than other tests or assignments that they had had in school. Also, twelfth graders who could choose a story had higher estimations of their performance on the assessment than did their counterparts who were assigned a story. On the other hand, no significant differences between choice and non-choice groups were observed in students' reports of their motivation for performing well on the assessment.

◆ *Patterns of Story Selection*

Despite some slight variations, the patterns of students' story selections were mostly similar across racial/ethnic groups at both grades 8 and 12, and across gender groups at grade 8. Among twelfth graders, however, males and females demonstrated strikingly different story preferences. Males were predominantly drawn to a story about a soldier and females were predominantly drawn to a story about a relationship.

◆ *Story Selection Criteria*

The most frequently reported basis for story selection in both grades was an affective or general evaluative criterion. Also, twelfth graders were more likely than eighth graders to select a story because it represented a particular genre.

◆ *Context Effects of Stories on Comprehension Questions*

Although identical questions were used to assess students' comprehension of each of the seven stories at each grade, there was evidence that many questions were more or less difficult to answer in conjunction with certain stories.

Background on the NAEP Reading Assessment

As educational theories and instructional approaches change over time to reflect evolving perceptions of how students learn and develop, concerns naturally arise about the assessment methods used to measure students' achievement. The emergence of an interactive, constructive theory of reading over the last two decades has not only brought about pedagogical reforms but has also called into question the traditional approaches of assessing reading development. In response, changes in how reading comprehension is measured can be observed in classrooms, in state-wide assessment initiatives, and in national large-scale assessment programs.

Reflecting these changing theories and practices, the National Assessment of Educational Progress (NAEP) reading assessment was redesigned in 1992 to include an increased emphasis on constructed-response questions and to involve students in reading authentic texts (materials selected from sources commonly available to students in and out of school). The assessment framework which provided the basis for developing the 1992 reading assessment views reading as a complex, interactive process between the reader, the text, and the context of the reading situation.[2] Furthermore, the processes and strategies used by readers to construct meaning from text are assumed to vary across texts and reading activities. As such, the framework specified that students should be assessed in reading for three different purposes: reading for literary experience, reading to gain information, and reading to perform a task.

The 1994 NAEP assessment of reading was conducted using two-thirds of the content from the 1992 assessment and new content that was developed from the same framework. Results from the 1994 assessment, as well as comparisons with results from the 1992 assessment are presented in *NAEP 1994 Reading Report Card for the Nation and the States*.[3]

Both the 1992 and 1994 NAEP reading assessments incorporated innovative tools and procedures for measuring reading comprehension that may be seen as responsive to the concerns of educators and researchers about more traditional testing approaches. For example, the use of authentic reading materials rather than passages that were written or abridged specifically for the assessment was viewed as creating a test situation which more closely replicated real-world reading tasks. Also, using a variety of texts representing different reading purposes rather than relying on a single type of text provided for a more comprehensive assessment. Emphasizing constructed responses to

---

[2] *Reading framework for the 1992 and 1994 National Assessment of Educational Progress*. (1994). National Assessment Governing Board. Washington, DC: U. S. Government Printing Office.

[3] Campbell, J. R., Donahue, P. L., Reese, C. M., & Phillips, G. W. (1996). *NAEP 1994 reading report card for the nation and the states*. National Center for Education Statistics. Washington, DC: U. S. Government Printing Office.

comprehension questions rather than relying primarily on multiple-choice formats provided opportunities for students to express fuller and more diverse interpretations based on their prior experiences and background knowledge.

Rationale for The NAEP Reader Study

Although the 1992 and 1994 NAEP reading assessments incorporated a number of innovations in measuring reading comprehension, many reading educators and researchers have voiced additional concerns about traditional assessment approaches. As the need to replicate tasks across students is paramount if comparisons between students' performance are to be made, standardized assessments of reading comprehension typically include a common set of reading materials and questions that are administered to all students participating in the assessment. Although fundamental principles of educational measurement require such a practice, it has been criticized by some within the field of reading as creating a situation in which test takers may lack the motivation and interest that support engagement and comprehension in more typical reading situations.[4]

The interaction of cognitive and affective processes has come to be viewed as an important aspect of readers' ability to comprehend texts. Some reading theorists have suggested that a reader's affective stance toward a text may play a critical role in the processes of comprehension.[5] Studies have shown that a positive attitude toward the reading task may increase the reader's attention, strategy use, and persistence.[6] Other studies indicate that the link between a reader's attitude and comprehension may be mediated by other variables, including the extent and relevance of prior knowledge, the task demands, and the context of the reading situation.[7]

---

[4] Levande, D. (1993). Standardized reading tests: Concerns, limitations, and alternatives. *Reading Improvement, 30*(2), 125-127.

[5] Mathewson, G. C. (1994). Model of attitude influence upon reading and learning to read. In R. B. Ruddell, M. R. Ruddell, & H. Singer (Eds.), *Theoretical models and processes of reading* (pp. 1131-1161). International Reading Association: Newark, DE.

Rosenblatt, L. M. (1994). The transactional theory of reading and writing. In R. B. Ruddell, M. R. Ruddell, & H. Singer (Eds.), *Theoretical models and processes of reading* (pp. 1057-1092). International Reading Association: Newark, DE.

[6] Alexander, P. A., Kulikowich, J. M., & Jetton, T. L. (1994). The role of subject-matter knowledge and interest in the processing of linear and nonlinear texts. *Review of Educational Research, 64*, 201-252.

Baldwin, R. S., Peleg-Bruckner, Z., & McClintock, A. H. (1985). Effects of topic interest and prior knowledge on reading comprehension. *Reading Research Quarterly, 20*(4), 497-504.

[7] Henk, W. A., & Homes, B. C. (1988). Effects of content-related attitude on the comprehension and retention of expository text. *Reading Psychology, 9*(3), 203-225.

Hollingsworth, P. M., & Reutzel, D. R. (1990). Prior knowledge, content-related attitude, reading comprehension: Testing Mathewson's affective model of reading. *Journal of Educational Research, 83*(4), 194-199.

The influence of affective process such as interest and motivation on reading comprehension and literacy development has become a central focus in numerous recent research studies and efforts to improve reading instruction.[8] It has been suggested that readers who are interested in the material and motivated to understand are more likely to demonstrate a level of engagement that promotes deeper levels of comprehension.[9] For example, readers who have interest in a text may more willingly engage in thoughtful consideration and be more apt to make personal connections with text ideas.

Often cited in the literature on engagement in reading is the body of research investigating the effects of intrinsic and extrinsic motivations on learning. Intrinsic motivations that are internal to the learner, such as interest, curiosity, and challenge, have been shown to promote and sustain higher levels of learning. Conversely, extrinsic motivations that are imposed externally, such as grades, recognition, and competition, may focus the learner on minimal levels of task completion.[10] Educators who seek to promote a life-long desire for reading in students and to provide students with the tools for succeeding at literacy tasks have come to recognize the importance of intrinsic motivation in classroom activities.

Increasingly, the growing knowledge base in literacy motivation and engagement has influenced school curriculum. For example, research indicating that student selection of tasks and materials can enhance learning attitudes and involvement have led to an emphasis on self-selected reading in many classrooms.[11] Recognizing that strong intrinsic motivation for reading is necessary to the student's development of strategies, such as summarizing and drawing inferences, many classrooms encourage such motivations as curiosity and involvement by allowing students to choose their own topics.

Providing students with a choice and giving students time to read books of their own choosing exemplify some of the effective strategies for literacy development that have become a part of instructional practice.[12] In addition, materials used for reading instruction are no longer limited to passages that were traditionally part of basal programs, passages that were usually written in a manner that controlled for vocabulary,

---

[8] Cramer, E., & Castle, M. (Eds.). (1994). *Fostering the love of reading: The affective domain in reading education.* International Reading Association: Newark, DE.

[9] Guthrie, J. T. (1996). Educational contexts for literacy engagement in literacy. *The Reading Teacher, 49*(6), 432-445.

Sweet, A. P., & Guthrie, J. T. (1996). How children's motivation relate to literacy development and instruction. *The Reading Teacher, 49*(8), 660-662.

[10] Deci, E. L., Vallerand, R. J., Pelletier, L. G., & Ryan, R. M. (1991). Motivation and education: The self-determination perspective. *Educational Psychologist, 26,* 325-346.

[11] Sweet, A. P. (1993, November). *Transforming ideas for teaching and learning to read.* Office of Educational Research and Improvement. U.S. Department of Education: Washington, DC.

[12] Raphael, T.E., & McMahon, S.I. (1994). Book club: An alternative framework for reading instruction. *The Reading Teacher, 48*(2), 102-116.

Turner, J., & Paris, S.G. (1995). How literacy tasks influence children's motivation for literacy. *The Reading Teacher, 48*(8), 662-673.

language, and topic. Instead, many teachers use a range of texts and text types in their instruction, giving students exposure to diverse reading materials and providing them opportunities to develop personal interests and preferences in reading.[13] By linking student's intrinsic motivations to curriculum activities, the classroom becomes a site of possibility for students to become engaged in and to further their own literacy development.

As the theory and practice of reading instruction evolve, it is important to consider the implication of these changing ideas on assessment procedures. Undoubtedly, the constraints of large-scale assessment do not allow for accommodating the infinite variety of interests and preferences of each individual participant. Indeed, as the assessment situation typically calls upon an extrinsic motivation of compliance, the degree to which a students' intrinsic motivation can be incited may be at least partially circumscribed. *The NAEP Reader* study was conceived as an examination of one concern voiced by educators and researchers -- the effects of choice on an assessment of reading comprehension. Set within the context of a large-scale assessment, the primary question addressed by this study is whether or not students perform differently on an assessment of reading comprehension when they are allowed to choose from a selection of texts rather than being given a particular text to read.

Design of The NAEP Reader Study

In order to examine the effect of choice, the *NAEP Reader* study was conducted with equivalent but distinct samples at each grade, differing only in whether or not they had a choice of which story to read. A nationally representative sample of 2,416 eighth graders and 2,100 twelfth graders was given a choice. These students, having received a collection of seven stories appropriate to their grade, were asked to select a story, to write a brief explanation of why they chose the story, and answer eleven constructed-response (open-ended) comprehension questions. The nationally representative samples that were assigned one of these same stories to read (i.e., one sample for each of the seven stories at each grade) ranged from 581 to 859 students at grade 8, and from 456 to 629 students at grade 12. The total number of students in the non-choice samples across all seven stories was 4,825 at grade 8 and 3,664 at grade 12. Students in these non-choice samples were asked to answer the same eleven comprehension questions for the assigned story as the choice sample answered in relation to a selected story.

---

[13] Hiebert, E. H. (1994). Becoming literate through authentic tasks: Evidence and adaptations. In R. B. Ruddell, M. R. Ruddell, & H. Singer (Eds.), *Theoretical models and processes of reading* (pp. 391-413). International Reading Association: Newark, DE.

Strickland, D. S. (1994/1995). Reinventing our literacy programs: Books, basics, balance. *The Reading Teacher, 48*(4), 294-302.

The collection of stories at each grade, entitled *The NAEP Reader*, comprised a variety of literary genres by both well- and lesser-known authors. The stories were drawn from sources appropriate to either grade 8 or grade 12 and were chosen for both their literary merit and cultural diversity. The length of the seven stories ranged approximately from 1,200 to 2,200 words at grade 8 and from 1,300 to 2,600 words at grade 12. While deemed comparable in difficulty by the committee of reading experts that oversaw the development of this study (see Appendix B), the stories covered distinctly different topics. Printed on the inside cover of each collection, very brief story summaries provided students with a hint about the plot or main character. On the facing page, the table of contents provided the authors' names. Thus, the collection resembled a literary text that students might encounter in school or in their reading experience. Figures 1 and 2 on the following pages present the story summaries which appeared at the beginning of *The NAEP Reader* for each grade.

Figure 1	**Eighth-Grade *NAEP Reader* Story Summaries**	

ALL SUMMER IN A DAY **by Ray Bradbury**	Here we have a group of children in a classroom on Venus, where the sun shines for only two hours once every seven years. For one of the children, however, the sun will not shine at all.
DREAM JOB **by Marjorie Weinman Sharmat**	Being a receptionist for a publishing company got boring awfully fast for sixteen-year-old Becky. It isn't a very exciting way for an aspiring writer to spend the summer. Then obnoxious Mr. REM pops into her life.
A DAY'S WAIT **by Ernest Hemingway**	Confusion surrounds the illness of a young boy who has resigned himself to dying until he learns the truth about his condition.
THE CIRCUIT **by Francisco Jimenez**	Picking fruit all day in the hot sun is hard work. But moving from town to town and starting life over again every few months can be even more difficult.
THE FULLER BRUSH MAN **by Gloria D. Miklowitz**	Selling brushes door to door after school is not an easy job for Donald. It is difficult to deal with the rejections, to handle the disappointments. But it is even more difficult for Donald to face his mother at home.
THE BOY WITH YELLOW EYES **by Gloria Gonzalez**	Norman was definitely weird. For one thing, all he ever did was read. Willie, on the other hand, was "a real boy" who especially loved baseball. What these two had in common came about only because a mysterious stranger came to town.
GREAT MOVES **by Sandy Asher**	Having the two most brilliant, most athletic, most handsome boys in the class fighting to take you to the dance might sound exciting to some girls. But while Jeff and Steve are fighting over Annie, no one has invited her best friend Brenda to the Valentine's Day dance.

Figure 2	**Twelfth-Grade *NAEP Reader* Story Summaries**	

LET ME PROMISE YOU **by Morley Callaghan**	In an attempt to salvage a failing relationship, Alice asks Georgie to visit her one winter evening after their break-up. As the evening progresses, their motivations for rekindling the relationship are revealed.
THE THIRD LEVEL **by Jack Finney**	Science rushes us into the future, yet the tools of science that have finally become part of our world are tame and represent access to a simpler past. In this science fiction story, the main character finds a new meaning for the word "nostalgia."
THE SNIPER **by Liam O'Flaherty**	Set against the backdrop of a bitter civil war in Dublin, Ireland at the turn of the century, a young man makes a startling discovery about the identity of his enemy.
CECIL RHODES AND THE SHARK **by Mark Twain**	For Cecil Rhodes, the catch of the day yields information that will change his life in a swift and calculated way.
THE LUMBER ROOM **by Saki**	The punishment Nicholas receives from his aunt turns into an afternoon of delight for him in a forbidden room and an ordeal for his aunt who falls into a rain water tank.
MURDER ON ST. VALENTINE'S DAY **by Mignon G. Eberhart**	Why would someone write a check for a face cream formula in lipstick on a heart-shaped handkerchief? Who murdered the inventor of the formula? These questions and others are answered in this murder mystery.
THE PORTRAIT **by Tomas Rivera**	A picture with a twist emerges when a dishonest portrait salesman crosses the path of Don Mateo — a man who is eager to preserve the memory of his deceased son.

In addition to the copy of *The NAEP Reader* appropriate for their grade, each student involved in the study received a booklet containing eleven comprehension questions. Of these eleven questions, eight were short-constructed response questions requiring a one or two sentence response and three were extended constructed-response questions requiring a more developed, reflective response of one or more paragraphs. Short constructed-response questions were scored as acceptable or unacceptable; extended constructed-response questions were scored according to a four-level rubric ranging from unsatisfactory to extensive. The assessment time was 50 minutes both for those students who were assigned a story and for those who were given a choice.

To accommodate students' choices and to allow for comparison of performance across the seven stories for students in both the choice and non-choice samples, the comprehension questions were composed generically so as to be applicable to any of the stories in the grade 8 or grade 12 *NAEP Reader*. For example, one of the questions asked students to describe the qualities of one of the main characters; another asked students to evaluate the appropriateness of the story's title. As these questions could be answered about any of the stories at each grade, all students participating in the assessment responded to the same set of questions. (The comprehension questions are presented Appendix A.)

For each grade, responses to *The NAEP Reader* comprehension questions were analyzed to determine the percentages of students responding in each of the categories specified by the scoring rubrics. The performance of the nationally representative student samples that were each assigned one of the seven stories was used to establish a scale. Item response theory (IRT) methods were used to produce the scale, which ranged from 0 to 100, with a mean of 50 and a standard deviation of 10. The performance of students who were allowed to choose a story was then analyzed using the same scale; thus, it is possible to report and compare students' performance in the choice and the non-choice samples on this scale.

An advantage of using IRT methods is that results for all students no matter which story they read are easily placed on the same scale. Three important assumptions were made in using this methodology. One is that each of the subsamples of students that were assigned a story to read is representative of the national student population. A second assumption is that for each story, each of the comprehension questions meant the same thing for students who selected the story and for students who were assigned the story. The third assumption is that the questions as answered in the context of each story all measure the same construct.

This Report

This report is comprised of three chapters, each focusing on a different aspect of the study. Chapter One presents findings related to the primary question of the study: Was student performance better when choice of stories was offered than when students were randomly assigned a story? Results are presented for the nation and by racial/ethnic and gender subgroups. In addition, students' perceptions of the assessment, including their motivation for performing well, are presented in this chapter. Chapter Two describes patterns of choices displayed by students who were allowed to select a story. Student selection patterns are presented for the nation, and by race/ethnicity and gender. Also in Chapter Two is a description of the selection criteria reported by students in making their story choices. Chapter Three examines how student performance on the generically worded questions varied in relation to different stories and presents sample student responses. The report concludes with a discussion of study results and issues related to study design and interpretations.

The average scale scores and percentages presented in this report are estimates because they are based on samples rather than the entire population. As such, the results are subject to a measure of uncertainty due to sampling error. In addition, measurement error contributes to the uncertainty of average scale scores reported for groups of students. The degree of uncertainty is reflected in the standard errors presented in parentheses along with the estimated average scores or percentages in tables and figures throughout this report.

The differences between scale scores or percentages discussed in the following chapters take into account the standard errors associated with the estimates. The comparisons are based on statistical tests that consider both the magnitude of the difference between the group average scores or percentages and the standard errors of those statistics. Throughout this report, differences are discussed only if they were determined to be statistically significant at the .05 level with appropriate adjustments for multiple comparisons.

Chapter 1

Choice vs. Non-Choice: Student Performance and Perceptions

This chapter presents results from *The NAEP Reader* study that address the effects of story selection on students' performance in a test of reading comprehension. The performance of students who were asked to choose one story to read from among seven stories is compared with the performance of students who were assigned one of the same seven stories to read. In addition, students' perceptions of the assessment and their motivation for completing the task in the choice and the non-choice situations are described and compared.

At both grades, students' average scores in the choice and non-choice samples are based on their responses to eleven comprehension questions about one of the seven stories in *The NAEP Reader*. The questions were the same regardless of the story chosen or assigned. In the choice sample, students were directed to select one of the seven stories to read; in the non-choice sample, students were told which of the seven stories to read. In both cases, the samples were selected to be nationally representative of students in grades 8 and 12. (See Appendix B for a more detailed discussion of sampling and administration.)

In order to describe students' performance on the comprehension questions, the non-choice sample at each grade was used to establish the parameters of a 0- to -100 scale, with a mean of 50 and a standard deviation of 10.[14] Thus, the average scale score for students on each story in the non-choice sample at both grades was set at 50. The performance of students in the choice sample was then placed on this scale, making it possible to compare their performance with that of their peers in the non-choice sample.

[14] Readers familiar with other NAEP reading reports may notice that the scale used for this study differs from the 0-to-500 scale used to report students' performance on the main NAEP reading assessment. Even though *The NAEP Reader* scale is not comparable to the main reading assessment scale in terms of content, readers should keep in mind that the units on this scale are about the same size as five units on the main assessment scale.

Table 1.1 presents average scores for eighth graders in both the choice and non-choice samples across all seven stories. Results are presented for the total sample as well as for racial/ethnic and gender subgroups of students. As shown in the table, the average score for the total sample of eighth-grade students who were allowed to choose a story was lower than the average score of all eighth graders who were not given a choice (49 compared to 50).

Table 1.1 **Choice vs. Non-Choice Average Scale Scores for Eighth-Grade Students**

	Average Scale Score		
	Choice		Non-Choice
Total	49.0(0.3)	<	50.0(0.2)
White	51.0(0.3)		52.0(0.3)
Black	44.2(0.5)		43.9(0.3)
Hispanic	44.1(0.6)		45.7(0.5)
Male	46.8(0.4)		47.3(0.3)
Female	51.6(0.3)	<	53.0(0.2)

< The value for the choice sample was significantly lower than the value for the non-choice sample.

The standard errors of the estimated scale scores appear in parentheses. It can be said with 95 percent certainty that for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1994 Reading Assessment.

Among eighth-grade students in each of the three racial/ethnic subgroups for which data are provided (White, Black, and Hispanic), no significant differences were observed between the performances of students who were given a choice and of students who were not given a choice. (The sample sizes for Asian, Pacific Islander, and Native American students were too small to allow for reporting the performance of these subgroups.) Similarly, the average scale score of male eighth graders in the choice sample did not differ significantly from that of males in the non-choice sample. However, among female eighth graders, a difference was observed: female students who were asked to choose a story had a lower average scale score than female students who were assigned a particular story to read (52 compared to 53).

Table 1.2 presents average scale scores for twelfth-grade students. For all twelfth graders and for students in each racial/ethnic and gender subgroup, no statistically significant differences were observed between the performance of students who were given a choice of stories and those students who were assigned a story to read.

Table 1.2 **Choice vs. Non-Choice Average Scale Scores for Twelfth-Grade Students**

	Average Scale Score	
	Choice	Non-Choice
Total	**49.6(0.3)**	**50.0(0.3)**
White	51.2(0.3)	51.4(0.3)
Black	44.8(0.5)	45.7(0.6)
Hispanic	46.9(0.7)	46.7(0.6)
Male	47.2(0.3)	47.3(0.3)
Female	52.3(0.4)	53.1(0.3)

The standard errors of the estimated scale scores appear in parentheses. It can be said with 95 percent certainty that for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1994 Reading Assessment.

One factor that may influence students' story selections is their reading ability. The possibility that students of different reading abilities may select different stories is one factor that can complicate making comparisons between students on an assessment that involves choice. As described earlier, the average scale score for students in the nationally representative samples who were assigned each story to read was set at 50 on a 0-to-100 scale, and the same scale was used to analyze the performance of students in the choice sample who selected to read the story. Consequently, it is possible to compare the average reading ability (as determined by performance on *The NAEP Reader* comprehension questions) of students in the choice sample who selected a particular story to the average reading ability of students who were assigned the same story in the non-choice sample.

Table 1.3 presents the average scale scores of students in the choice and non-choice samples at both grades 8 and 12. The results indicate that for each story at both grades, the average score of students who selected to read a story did not differ significantly from the average score of 50 for students who were assigned the same story (NOTE: For each story in the non-choice sample, the mean score was set at 50. See Appendix B for details.) If a significant difference in average scale scores had been observed between students in the choice and non-choice groups for any story, it may have provided some evidence that students of a certain ability level were more or less likely to select that story. However, the results of this analysis provide no evidence that this occurred.

Table 1.3 **Average Scale Scores of Eighth- and Twelfth-Grade Students in the Choice and Non-Choice Samples by Story**

THE NATION'S REPORT CARD naep

	Story 1	Story 2	Story 3	Story 4	Story 5	Story 6	Story 7
Grade 8							
Choice	48.4 (0.4)	48.6 (0.7)	51.1 (0.7)	50.8 (1.1)	48.2 (1.6)	48.8 (0.7)	47.7 (0.9)
***Non-Choice**	50.0 (0.5)	50.0 (0.5)	50.0 (0.5)	50.0 (0.5)	50.0 (0.5)	50.0 (0.5)	50.0 (0.5)
Grade 12							
Choice	50.4 (0.6)	50.0 (0.9)	49.0 (0.6)	48.7 (1.1)	50.1 (1.6)	50.2 (0.5)	47.3 (1.3)
***Non-Choice**	50.0 (0.6)	50.0 (0.5)	50.0 0.6)	50.0 (0.6)	50.0 (0.7)	50.0 (0.5)	50.0 (0.6)

*For each story in the non-choice sample, the mean score was set at 50. See Appendix B for details.

The standard errors of the estimated scale scores appear in parentheses. It can be said with 95 percent certainty that for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1994 Reading Assessment.

One question that may be asked about students' performance in the choice situation is whether or not the process of selecting a story resulted in less time that could be devoted to answering the comprehension questions. Students in both the choice and non-choice samples were given the same 50-minute time period to complete the assessment. No additional time was allotted for students to make a selection and to explain why they chose a story.

One measure of the speededness, or adequacy of time allotted for the assessment, is the percentage of students who attempt all of the questions presented to them. The analysis of student performance distinguished between missing responses at the end of the set of questions and missing responses prior to the last observed response. Missing responses before the last observed response were considered intentional omissions. Missing responses at the end of the set of questions were considered "not reached."

Table 1.4 presents the percentages of students in both samples at each grade who did not reach the last two questions in the assessment, providing some indication of whether or not the amount of time given to students was adequate for completing the assessment. The last two questions are examined for the purposes of this analyses rather than only the last question since it is unknown if students who did not answer the last question, but answered the next to last question, did not have enough time to respond to the last question or simply chose to omit it. Students who are included in the "not reached" percentage for the next to last question did not respond to that question or the last question. Thus, there is stronger evidence that these students did not have enough time to complete the entire set of questions. Percentages are presented for each story in the choice and the non-choice situation.

Table 1.4 **Percentages of Students Not Reaching the Last Two Comprehension Questions**

	Percentage of Students			
	Grade 8		Grade 12	
	Choice	Non-Choice	Choice	Non-Choice
Story 1	7 (1.2)	6 (1.2)	3 (0.8) <	9 (1.5)
Story 2	6 (1.3)	8 (1.4)	3 (1.7)	7 (1.6)
Story 3	4 (1.3)	3 (0.9)	4 (0.9)	5 (1.2)
Story 4	6 (2.9)	5 (1.2)	6 (2.5)	4 (1.4)
Story 5	13 (7.6)	6 (1.1)	11 (4.2)	9 (1.9)
Story 6	12 (2.1)	9 (1.5)	10 (1.3)	9 (1.6)
Story 7	14 (3.6)	14 (1.9)	11 (3.9)	6 (1.8)

< The value for the choice sample was significantly lower than the value for the non-choice sample.

The standard errors of the estimated percentages appear in parentheses. It can be said with 95 percent certainty that for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1994 Reading Assessment.

BEST COPY AVAILABLE

One significant difference was observed between the percentages of student not reaching the last two comprehension question in the choice and the non-choice samples. For story 1 at grade 12, the percentage of students in the choice sample who did not reach the last two questions was lower than the percentage of students in the non-choice sample. These data provide some indication that, in comparison to their peers who were assigned a story, students who were allowed to choose a story were not significantly disadvantaged in the amount of time they had to read the story and answer the comprehension questions.

Students' Perceptions of the Assessment and Their Motivation

Factors that may contribute to how well students perform on an assessment of reading comprehension include their perception of the difficulty of the task and their motivation for doing well on the assessment. Examining how hard students perceived the assessment to be, how well they thought they performed on the assessment, how hard they tried, and how important they felt it was to perform well on the assessment can further illuminate the effects of choice in an assessment of reading comprehension.

As a part of *The NAEP Reader* study, students were asked a series of questions about their perceptions of the assessment and their motivation for performing well. Table 1.5 presents students' responses in the choice and non-choice samples at both grades to the first question: "About how many questions do you think you got right on the reading test you just took?"

Table 1.5 **Students' Perceptions of How Many Questions They Got Right**

	Percentage of Students			
	All	More Than Half	About Half	Less Than Half
Grade 8				
Choice	31 (1.3)	38 (1.3)	23 (1.1)	8 (0.9)
Non-Choice	29 (0.8)	39 (0.8)	24 (0.8)	8 (0.5)
Grade 12				
Choice	38 (1.1)>	33 (1.1)	20 (0.9) <	8 (0.7)
Non-Choice	32 (1.0)	34 (0.9)	23 (0.7)	11 (0.7)

<The value for the choice sample was significantly lower (>higher) than the value for the non-choice sample.

The standard errors of the estimated percentages appear in parentheses. It can be said with 95 percent certainty that for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1994 Reading Assessment.

At grade 8, no significant difference was observed between the choice and non-choice samples in students' perceptions of how many questions they got right in the assessment. At grade 12, however, there was evidence that students who were allowed to choose a story thought they got a larger proportion of the questions right than did students who were not given a choice. The percentage of twelfth graders who thought that they got all of the questions right was higher in the choice sample than in the non-choice sample. Correspondingly, the percentage of students in the choice sample who thought they got only about half of the questions right was smaller than the percentage of students in the non-choice sample who reported the same estimation of their performance.

Table 1.6 presents students' responses to a related question: "How hard was this test compared to most other reading tests or assignments you have had in school this year which asked you to answer questions about something you read?"

Table 1.6 **Students' Perceptions of How Hard This Test Was Compared to Other Reading Tests or Assignments**

	Percentages of Students			
	Much Harder	Harder	About as Hard	Easier
Grade 8				
Choice	6 (0.7)	16 (0.9)	37 (1.0)	40 (1.2) >
Non-Choice	7 (0.4)	18 (0.6)	38 (1.0)	36 (0.9)
Grade 12				
Choice	2 (0.3)<	8 (0.6) <	37 (1.0)	52 (1.0) >
Non-Choice	4 (0.4)	11 (0.6)	38 (0.8)	47 (1.0)

< The value for the choice sample was significantly lower (> higher) than the value for the non-choice sample. The standard errors of the estimated percentages appear in parentheses. It can be said with 95 percent certainty that for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1994 Reading Assessment.

At both grades 8 and 12, students who were given a choice of story were more likely to rate this assessment as easier in comparison to other reading tests or assignments they had had in school. In addition, at grade 12 the percentages of students in the choice sample who described this assessment as either "much harder" or "harder" were lower than the percentages of students in the non-choice sample who made the same comparisons with other tests and assignments. It cannot be said with certainty from these data whether it was being able to select a story that led students to perceive the task as being easier, or whether the story itself was perceived to be easier. That is, it is possible that students who were allowed to select a story chose one that they perceived to be easier than the others.

Table 1.7 presents students' responses to a question about the extent to which they were motivated to perform well: "How hard did you try on this test compared to how hard you tried on most other reading tests or assignments you have taken this year in school?" At both grades, there was no significant difference between choice and non-choice samples in students' reports of how hard they tried on this test.

Table 1.7 **Students' Reports About How Hard They Tried Compared to Other Reading Tests or Assignments**

	Percentage of Students			
	Much Harder	Harder	About as Hard	Not as Hard
Grade 8				
Choice	14 (0.8)	19 (1.0)	48 (1.1)	19 (1.3)
Non-Choice	13 (0.6)	18 (0.6)	50 (0.8)	18 (0.8)
Grade 12				
Choice	4 (0.5)	9 (0.7)	50 (1.2)	37 (1.2)
Non-Choice	5 (0.4)	9 (0.5)	50 (0.8)	36 (1.0)

The standard errors of the estimated percentages appear in parentheses. It can be said with 95 percent certainty that for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1994 Reading Assessment.

Table 1.8 presents students' responses to a related question about their motivation for performing well on the assessment: "How important was it to you to do well on this reading test?" Similar to students' reports to the previous question, there were no significant differences between the choice and non-choice samples at either grade in students' reports of how important it was to them to do well on the assessment.

Table 1.8 **Students' Reports of How Important It Was to Them to Do Well**

	Percentage of Students			
	Very Important	Important	Somewhat Important	Not Important
Grade 8				
Choice	30 (1.4)	32 (1.0)	24 (1.3)	14 (0.9)
Non-Choice	28 (0.9)	32 (0.8)	26 (0.8)	14 (0.7)
Grade 12				
Choice	13 (0.8)	29 (1.1)	32 (1.2)	27 (1.2)
Non-Choice	12 (0.7)	28 (1.0)	34 (0.9)	25 (1.1)

The standard errors of the estimated percentages appear in parentheses. It can be said with 95 percent certainty that for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1994 Reading Assessment.

Summary and Discussion

The findings presented in this chapter suggest that allowing students to choose a story from a collection of stories in an assessment of reading comprehension did little to improve their performance. At grade 12, no significant difference was observed between the average scale scores of students who were allowed to choose a story and the performance of students who were assigned one of the stories to read. Among eighth graders, however, small but statistically significant differences were observed between the choice and the non-choice samples. The average scale score of eighth graders who were allowed to choose a story was *lower* than the average scale score of their counterparts who were assigned a story to read. The difference between the average performance of the two samples was 1 scale point on a 0-to-100 scale. When results for subgroups of eighth-grade students were examined, only one statistically significant difference was observed. Among female eighth graders, a lower average scale score was attained by students who were allowed to choose a story compared to students who were assigned a story.

As described in the following chapter, students who were allowed to select a story demonstrated a variety of choice patterns and employed various selection criteria. Assuming that students selected stories based on interest and personal preferences, it may be questioned why the opportunity to do so did not result in higher scores, particularly since there were some differences in their perceptions of the assessment. Students who were given a choice were more likely than students who were not given a choice to rate the assessment as easier than other reading tests or assignments. This difference in perception was observed at both grades 8 and 12. Additionally, twelfth graders in the choice sample had a higher estimation of their performance on the comprehension questions than did twelfth graders in the non-choice sample.

Students who were allowed to select a story perceived the assessment to be easier than did students who were assigned a story. However, students' responses to questions about how hard they tried and how important it was for them to perform well reflected no differences between the motivation of students in the choice or non-choice samples. These results suggest that allowing students to select a story to read for an assessment may affect their perceptions of the task and of their performance, perhaps making the test seem less difficult and increasing students' confidence. While this may be viewed as a more positive affective state for taking a test, it did not appear to translate into increased motivation or, ultimately, increased performance.

Chapter 2

Patterns of Students' Story Selections

This chapter describes the selections made by students who were allowed to choose a story to read in *The NAEP Reader* study. Percentages of eighth- and twelfth-grade students who selected each of the stories, as well as patterns of selection by race/ethnicity, and by gender, are presented. Before responding to the comprehension questions, students participating in *The NAEP Reader* study who had a choice as to which story to read were asked to explain why they selected that story. The results of students' responses about why they selected a particular story conclude this chapter. Examining the selection patterns of different groups of students and the reason they reported for making their selections not only provides insight into their literary preferences, but also provides another perspective on their overall performance. In order to better understand the interests and abilities reflected by the selections students made, descriptions of the stories in *The NAEP Reader* for each grade are provided on the following pages.

The NAEP Reader — Grade 8

◆ Story 1: *All Summer In A Day* by Ray Bradbury

The situation in this science fiction story about life on a planet where the sun appears for only two hours every seven years reflects an experience common to growing up — the difficulty of being different from other children and the careless cruelty children inflict on those different from themselves. Margot is a quiet child who, having moved to Venus from Earth, remembers what the sun looks and feels like. Resenting Margot's aloofness and memory of the sun, the children exclude her. On the day the sun is to shine, they lock her in a closet and forget about her while they enjoy the light and warmth. Only when it begins to rain again do they remember Margot and feel ashamed for what they have done.

◆ Story 2: *Dream Job* by Marjorie Weinman Sharmat

The narrator of this story is a sixteen-year-old named Becky who is working as a receptionist in a publishing company for the summer. Bored, and resentful that her major job duty is to smile, this would-be professional writer escapes the constraints of her job via her unconstrained imagination. Oscillating between adolescent arrogance and angst, the colloquial tone sustains a narrative that is both realistic and fantastic. "Dream Job" is the dream that Becky has while sleeping on the job at the publishing company.

◆ Story 3: *A Day's Wait* by Ernest Hemingway

In the simple sentences characteristic of Hemingway's style and developed mainly through dialogue, this story relates the incidents of a single day and the interchange between a nine-year-old boy and his father. At the start, the boy has a high fever and is diagnosed by the doctor as having influenza. Confined to bed, and confused about the degree of his temperature, the boy refuses visitors and does not listen as his father reads aloud to him. At the end of the story the significance of the title is evident as the boy reveals that he has been waiting all day to die.

◆ **Story 4:** *The Circuit* by Francisco Jimenez

The stark realities of migrant farmers' existence are vividly rendered in this story that Panchito, the second oldest son, relates about his family. Mundane details and simple conversations accumulate to convey the hard work and impoverished conditions of itinerant life. More than the work, however, what Panchito finds most difficult is having to pack up and move every few months. The story begins and ends with the image of packed boxes that represent not only the rootlessness of itinerant life, but also enclosure and the lack of opportunities.

◆ **Story 5:** *The Fuller Brush Man* by Gloria D. Miklowitz

This story relates a teenager's resistance to accepting that his mother is ill and his refusal to visit her. Donald, the main character, works after school selling door to door, not only to earn money for college but also to avoid going home. He has withdrawn inside himself in response to his mother's illness as he withdraws when customers close the door in his face. His own memories, the urgings of others, and his fear that he might be too late force Donald to visit his mother's room; his job as a Fuller Brush man provides him with an identity and a fixed smile to help him get through her door.

◆ **Story 6:** *The Boy With Yellow Eyes* by Gloria Gonzalez

An aura of local legend pervades this narration of an unlikely alliance between two very different boys in an adventure that leads to lunch with the vice-president of the United States and to a lifelong friendship. When quiet, bookish Norman and daring, athletic Willie encounter a "stranger" out by abandoned railroad cars, it is Norman who is able to read the code, recognize the stranger as a Nazi spy, and tackle him while Willie stands paralyzed with fear before bringing his baseball bat to the rescue.

◆ **Story 7:** *Great Moves* by Sandy Asher

In this story, Annie and Brenda, who have been best friends since elementary school, experience conflicting emotions when two boys compete to take one, then the other, of them to the high school dance. Both boys are local heroes, stars in athletics, school government, and their studies, but they are not so bright when it comes to getting a date for the dance. Watching their so-called admirers fight it out, Annie and Brenda realize how little they figure in the competition. Their friendship is renewed as they learn that sometimes the greatest move is just standing your ground and saying no.

The NAEP Reader — Grade 12

◆ Story 1: *Let Me Promise You* by Morley Callaghan

This story has only two characters whose conversation and unspoken thoughts constitute much of the story's action. Alice has asked her old boyfriend Georgie to visit her and he arrives despite the drizzling weather. His reaction to the extravagant gifts she has bought him for his birthday reveal the ambivalence and confusion of Georgie's feelings for Alice. He simultaneously does and does not want them. Alice is equally confused in her emotions, poised between affection for Georgie and resentment of his rejection of her and her gifts. Only when Georgie sees Alice on her knees does he realize his affection for her, as he had realized how much he wanted the watch once he saw it smashed to pieces.

◆ Story 2: *The Third Level* by Jack Finney

The adult narrator of this science-fiction fantasy lives in New York City and longs for both the rural simplicity of his childhood and the innocence of life before the World Wars. Using realistic details, he recounts happening upon an entry to the past while lost in Grand Central Station and being dissuaded from pursuing this "escape" by both his wife and his psychiatrist. His nostalgia, however, is contagious; though he hadn't totally convinced himself of the reality of the past, he had unwittingly convinced his psychiatrist. A letter postmarked 1894 that mysteriously appears in his grandfather's stamp collection confirms his belief in the power of his imagination and convinces him to pursue his dream.

◆ Story 3: *The Sniper* by Liam O'Flaherty

Preceded by a brief summary that provides the historical context, this story relates the events of one night as experienced by a Republican soldier fighting to free Ireland from British rule during the Civil War of 1922-1923. Comprised of a straightforward description of a single character's thoughts and actions, the story contains no dialogue. A strong element of suspense is sustained throughout as the wounded soldier must outwit the enemy he cannot see in the darkness of night on the opposite rooftop. When he succeeds and his enemy falls, the surprising and tragic ending is foreshadowed by the sniper's sudden feeling of remorse. The horrors of war and the tragedy of a country divided against itself are effectively represented by the sniper's realization that he has killed his own brother.

◆ **Story 4:** *Cecil Rhodes and the Shark* by Mark Twain

Announcing at the start, "I have a tale to tell, which has not as yet been in print," the narrator establishes a sense of verisimilitude and then goes on to relate rather extraordinary events. Using 19th-century idioms and diction peculiar to fishermen, he tells of a young man (Cecil Rhodes) down on his luck wandering in Sydney in the year 1870. The reversal of his fortune involves landing a shark, robbing its belly of a journal and a London newspaper, and using the information therein to convince a merchant to loan him thousands of pounds for a business venture. Much of the story consists of the conversation between Rhodes and the merchant. Impressed by his confidence but doubtful that he could have a newspaper only ten days old out of London, the merchant oscillates between thinking Rhodes either remarkable or crazy. Swayed by Rhodes' financial calculations and convinced by actually seeing the London newspaper, the merchant classifies Rhodes as remarkable and loans him the cash to secure his first fortune.

◆ **Story 5:** *The Lumber Room* by Saki

The omniscient narrator of this story is in sympathy with the protagonist, Nicholas, a precocious child who outwits the adults around him. Unduly punished by his cousins' aunt for a harmless prank, Nicholas' restriction to home provides him with an experience that opens to him the world of fiction. When he unlocks the door to the lumber room, where unused furnishings are stored, he studies the story portrayed in a tapestry and learns the pleasures of interpretation. Meanwhile, the narrow-minded aunt is physically restricted when she slips down the rainwater tank in the garden. Precocious Nicholas questions the prisoner, thus revealing the silly ways adults attempt to control children, then leaves her to be rescued by someone else. The silence at dinner that evening is difficult for everyone at the table except Nicholas who is at ease in his imagination contemplating the possible escape of the huntsman from the wolves in the tapestry he had studied that afternoon.

◆ **Story 6:** *Murder on St. Valentine's Day* by Mignon G. Eberhart

When James Wickwire, the banker who advises Clarissa Hartridge about her finances, sees that she has written a check for a large amount of money on a heart-shaped handkerchief, his belief in her financial astuteness is shaken. He suspects that she has fallen in love, for the recipient of the check is young and handsome. His suspicion seems confirmed when the young man, whom he knows has a fiancee, is found dead in Clarissa's library with a key to her house in his pocket. However, the conclusion of a conversation with Clarissa, from which the reader is excluded, leads Wickwire to doubt her culpability. When her housemate, Miss Gray, returns from a seemingly innocent

shopping excursion — without her house key — Wickwire uncovers both the motive and the murderer. Missing from the dead man's pockets was a formula for which Clarissa had written the check and this is found concealed in Miss Gray's purchases. At story's end, Clarissa is exonerated not only from murder but also from financial irresponsibility. She uses the handkerchief as a trademark in a marketing venture that makes a lot of money and Wickwire realizes it is *she* who should be giving *him* financial advice.

◆ **Story 7:** *The Portrait* by Tomas Rivera

The son of a friend of Don Mateo tells of the day the portrait salesmen in white shirts came to San Antonio to take advantage of the recently paid Hispanic workers. This story of their deception and of Don Mateo's subsequent retribution is rendered almost totally in dialogue. The slick language of the salesman's pitch manipulates the emotions of Don Mateo and his wife. Promised a life-like portrait, they hand over their hard-earned cash and their only photograph of the son they lost in the Korean war. When weeks pass and no portrait arrives, Don Mateo tracks down the swindler in San Antonio and scares him into producing the promised portrait in just three days.

Overall Pattern of Students' Selections of Stories in The NAEP Reader

The overall percentages of students at grades 8 and 12 that selected each story are presented in Table 2.1. While the largest percentage of students selecting a particular story was 27 percent at both grades, eighth-grade students demonstrated greater variety in their selections than did twelfth graders. Five of the seven stories in the eighth-grade *Reader* were selected by at least 10 percent of the students. While the highest percentage of eighth graders, 27 percent, selected Story 1, approximately equal percentages of students chose the 2nd, 3rd, and 6th stories (18 to 19 percent). At grade twelve the distribution of students' selection was less varied. While the highest percentage of twelfth graders, 27 percent, chose to read Story 3, this was not significantly higher than the percentages that selected Story 1 and Story 6 (25 and 23 percent). The remaining four stories were selected by less than 10 percent of the twelfth graders.

Table 2.1 **Percentages of Students Selecting Each Story in *The NAEP Reader***

THE NATION'S REPORT CARD

Grade 8

Story 1 *All Summer In a Day*	Story 2 *Dream Job*	Story 3 *A Day's Wait*	Story 4 *The Circuit*	Story 5 *The Fuller Brush Man*	Story 6 *The Boy with Yellow Eyes*	Story 7 *Great Moves*
27 (1.3)	19 (1.0)	18 (0.9)	4 (0.5)	3 (0.5)	18 (1.1)	10 (0.9)

Grade 12

Story 1 *Let Me Promise You*	Story 2 *The Third Level*	Story 3 *The Sniper*	Story 4 *Cecil Rhodes and the Shark*	Story 5 *The Lumber Room*	Story 6 *Murder on St. Valentine's Day*	Story 7 *The Portrait*
25 (1.0)	8 (0.7)	27 (1.1)	7 (0.6)	3 (0:4)	23 (1.2)	6 (0.6)

The standard errors of the estimated percentages appear in parentheses. It can be said with 95 percent certainty that for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1994 Reading Assessment.

Pattern of Students' Story Selections by Race/Ethnicity

Table 2.2 presents the percentages of eighth-grade students by race/ethnicity who selected each story in *The NAEP Reader*. Little variation in story selection was evident among the racial/ethnic subgroups. At grade 8, the largest percentage of White, Black, and Hispanic students chose story 1, *All Summer In a Day* by Ray Bradbury. While each of the stories at grade 8 was chosen by relatively similar percentages of White, Black, and Hispanic students, there were two exceptions. The percentage of White students who chose story 3, *A Day's Wait* by Ernest Hemingway, was significantly higher than the percentages of both Black and Hispanic students who chose that story. While the percentage of White eighth graders who chose story 4, *The Circuit* by Francisco Jimenez, was significantly higher than the percentage of Black students, only a low percentage of both groups chose that story.

Table 2.2 **Percentages of Eighth-Grade Students Selecting Each Story in *The NAEP Reader* by Race/Ethnicity**

THE NATION'S REPORT CARD

	Story 1 *All Summer In a Day*	Story 2 *Dream Job*	Story 3 *A Day's Wait*	Story 4 *The Circuit*	Story 5 *The Fuller Brush Man*	Story 6 *The Boy with Yellow Eyes*	Story 7 *Great Moves*
White	26 (1.6)	18 (1.2)	20 (1.2)	5 (0.6)	3 (0.5)	18 (1.3)	10 (1.1)
Black	30 (2.9)	22 (2.4)	11 (1.5)	2 (0.8)	4 (1.5)	22 (2.6)	10 (1.6)
Hispanic	30 (2.9)	20 (3.8)	14 (1.8)	4 (1.3)	3 (1.1)	17 (2.4)	12 (1.7)

The standard errors of the estimated percentages appear in parentheses. It can be said with 95 percent certainty that for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1994 Reading Assessment.

Table 2.3 presents the percentages of twelfth graders by race/ethnicity who selected each story in *The NAEP Reader*. At grade 12, the distribution of White, Black, and Hispanic students' selections displayed similar patterns across the stories. The vast majority of each racial/ethnic subgroup chose story 1, story 3 or story 6; and the lowest percentage of each subgroup chose story 5. The only significant difference in selections between racial/ethnic subgroups at grade 12 was with story 3, *The Sniper* by Liam O'Flaherty, where a higher percentage of White than Black students chose to read the story.

Table 2.3 **Percentages of Twelfth-Grade Students Selecting Each Story in *The NAEP Reader* by Race/Ethnicity**

THE NATION'S REPORT CARD naep

	Story 1 *Let Me Promise You*	*Story 2* *The Third Level*	*Story 3* *The Sniper*	*Story 4* *Cecil Rhodes and the Shark*	*Story 5* *The Lumber Room*	*Story 6* *Murder on St. Valentine's Day*	*Story 7* *The Portrait*
White	24 (1.4)	8 (0.9)	29 (1.2)	8 (0.9)	4 (0.4)	22 (1.6)	5 (0.7)
Black	30 (2.6)	9 (1.4)	20 (2.7)	4 (0.9)	2 (0.9)	28 (2.5)	8 (1.4)
Hispanic	28 (3.1)	6 (2.1)	25 (2.7)	5 (1.1)	2 (0.8)	24 (2.9)	9 (1.6)

The standard errors of the estimated percentages appear in parentheses. It can be said with 95 percent certainty that for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1994 Reading Assessment.

Pattern of Students' Story Selections by Gender

As displayed in Table 2.4, eighth-grade male and female students demonstrated similar patterns in their selections of stories. The most frequently selected story for both males and females (26 and 29 percent respectively) was story 1, a science fiction tale involving a classroom on planet Venus. The remaining selections shared a fairly similar pattern with the next most frequently selected stories — story 2, story 3, and story 6 — being chosen by 16 to 18 percent of males and 17 to 22 percent of females. While it might be assumed that males will choose stories with male protagonists and that females will choose stories with female protagonists, at grade 8 gender appeared to have little effect on students' selections. For example, 20 percent of females and 17 percent of males selected *The Boy with Yellow Eyes*, a story in which both main characters are males.

Table 2.4 **Percentages of Eighth-Grade Students Selecting Each Story in *The NAEP Reader* by Gender**

THE NATION'S REPORT CARD

	Story 1 *All Summer In a Day*	*Story 2* *Dream Job*	*Story 3* *A Day's Wait*	*Story 4* *The Circuit*	*Story 5* *The Fuller Brush Man*	*Story 6* *The Boy with Yellow Eyes*	*Story 7* *Great Moves*
Males	26 (1.7)	16 (1.3)	18 (1.3)	6 (0.9)	5 (0.8)	17 (1.4)	12 (1.2)
Females	29 (1.6)	22 (1.5)	17 (1.2)	2 (0.5)	1 (0.4)	20 (2.5)	9 (0.9)

The standard errors of the estimated percentages appear in parentheses. It can be said with 95 percent certainty that for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1994 Reading Assessment.

As shown in Table 2.5, twelfth-grade male and female students differed from each other in their selection patterns. While *The Sniper* was the most popular selection among male twelfth graders (45 percent), only 8 percent of female twelfth graders made this selection. Similarly, the most frequently selected story among female twelfth graders (43 percent) was story 1, *Let me Promise You*. In contrast, only 9 percent of males chose to read this story about a relationship. The second most frequently selected story by both male and female twelfth graders was story 6, *Murder on St. Valentine's Day*. However, the percentage of females (31 percent) who selected this story that combines murder and romance in a mystery was significantly higher than the percentage of males (16 percent) who chose to read the story. It may be of interest that the gender of a story's author appeared to have little effect on either males' or females' story selections at both grades.

Table 2.5 **Percentages of Twelfth-Grade Students Selecting Each Story in *The NAEP Reader* by Gender**

THE NATION'S REPORT CARD

	Story 1 *Let Me Promise You*	*Story 2* *The Third Level*	*Story 3* *The Sniper*	*Story 4* *Cecil Rhodes and the Shark*	*Story 5* *The Lumber Room*	*Story 6* *Murder on St. Valentine's Day*	*Story 7* *The Portrait*
Males	9 (0.8)	11 (1.1)	45 (1.6)	10 (0.9)	4 (0.6)	16 (1.1)	5 (0.7)
Females	43 (1.8)	5 (0.8)	8 (1.0)	4 (0.7)	3 (0.5)	31 (2.0)	7 (0.9)

The standard errors of the estimated percentages appear in parentheses. It can be said with 95 percent certainty that for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1994 Reading Assessment.

Students' Criteria for Selecting Stories

Those students participating in *The NAEP Reader* study who were given a choice as to which story to read were asked to explain the basis for their selection. Their explanations were classified according to a coding scheme that had been developed using student response data from the 1992 *NAEP Reader* study, which served as a pilot for the 1994 study. (See Appendix B for a description of the 1992 pilot.) Responses were coded for the selection criteria presented in Figure 2.1. If a student's response included more than one selection criterion, for instance if it mentioned both that it was a catchy title and that they recognized the author's name, the response was coded for both selection criteria.

Figure 2.1	Coding Categories for Describing Students' Story Selection Criteria	

Affective or General Evaluative Comment	Student's response indicates the selection was influenced by an anticipated emotional reaction, such as being scared or saddened, or by a general sense that the story would prove interesting or exciting.
Title/Table of Contents	Student's response indicates that the selection was influenced by a reaction to the title or by relying on the table of contents.
Topic	Student's response indicates that the selection was influenced by the subject matter.
Genre	Student's response indicates that the selection was influenced by an awareness or appreciation of the particular genre represented by the story.
Personal Identification	Student's response indicates that the selection was influenced by some sense of identification with the character or plot of the story, or by having connected an aspect of the story to something in his or her own life.
Position in Book/Length or Difficulty of Story	Student's response indicates that the selection was influenced by its position in *The NAEP Reader*, or by some aspect of the text, such as length or vocabulary, that made it seem either easier or more difficult than the other stories.
Character/Plot/Setting	Student's response indicates that a specific story element influenced their selection, for instance, wanting to read about a male main character, about a relationship, or about an incident that takes place on another planet.
Author	Student's response indicates that the selection was influenced by some aspect of authorship, such as having read other stories by the same author, having heard of the author, being drawn to the author's name in itself, or for reasons of author's gender or ethnicity.
Summary/Skimmed Story	Student's response indicates that the selection was influenced by having read the summary provided in *The NAEP Reader*, or by having skimmed or read portions of some or all of the stories.
Knew Story	Student's response indicates that the selection was influenced by having seen a movie based on the story, having read the story before, or knowing someone who read the story.
No Reason	Student's response indicates that the selection was made randomly with no operative criterion; for example, just having opened to a story and starting to read.

Table 2.6 presents the distribution of students' responses to the question which asked them to explain why they chose their story. The percentages provided in the table represent the percent of all the criteria indicated by students, providing some indication of the emphases placed by students on different selection criteria. For example, 40 percent of all the selection criteria reported by eighth-grade students (many students indicated more than one criterion) were considered to be an *Affective or General Evaluative Comment*.

Table 2.6 **Students' Reports of Story Selection Criteria**

Selection Criteria	Percentage of All Criteria Grade 8		Grade 12
Affective or General Evaluative Comment	40 (0.9)	>	30 (0.6)
Title/Table of Contents	17 (0.6)		18 (0.7)
Topic	6 (0.4)	<	8 (0.5)
Genre	3 (0.3)	<	15 (0.7)
Personal Identification	10 (0.7)	>	6 (0.5)
Position in Book/Length or Difficulty of Story	6 (0.5)		6 (0.6)
Character/Plot/Setting	6 (0.4)	>	3 (0.4)
Author	3 (0.4)		4 (0.3)
Summary/Skimmed Story	3 (0.3)	<	5 (0.4)
Knew Story	3 (0.3)		2 (0.3)
No Reason	3 (0.4)		2 (0.3)

<The value for eighth-grade students was significantly lower (> higher) than the value for twelfth-grade students.

The standard errors of the estimated percentages appear in parentheses. It can be said with 95 percent certainty that for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1994 Reading Assessment.

Forty percent of eighth grade criteria and 30 percent of twelfth grade criteria were categorized as *Affective or General Evaluative Comment*. The next most frequently reported criteria, 17 percent at grade 8 and 18 percent at grade 12, was *Title* or *Table of Contents*. Eighth graders were more likely than twelfth graders to base their selections on *Personal Identification* and *Character/Plot/Setting*, while twelfth graders were more likely to consider *Topic* and *Genre*. The difference between grades in the use of genre as a criterion was particularly notable: whereas 15 percent of twelfth graders selected which story to read on the basis of genre, only 3 percent of eighth graders indicated that genre served as the basis for their choice. Also, twelfth-grade students more frequently reported that they used the story summaries or skimmed the stories to make their selections.

Summary and Discussion

As demonstrated by the results presented in this chapter, when students were allowed to choose a story to read from among seven for an assessment of reading comprehension, variations in their selections were apparent. At grade 8, students displayed more variation than twelfth graders in that five of the seven stories were selected by at least 10 percent of eighth graders. At grade 12, only three stories were selected by at least 10 percent of students. It is not known from these data, however, if this represents a greater variation in the preferences of eighth graders or if the stories contained in the eighth-grade *Reader* presented a greater range of appeal than those contained in the twelfth-grade *Reader*.

At both grades, few differences were observed between the choices made by students in different racial/ethnic groups. Gender effects on story selections were also minimal at grade 8. However, a very clear difference was observed at grade 12 between the predominant stories selected by males and females. Twelfth grade female students were overwhelming more likely than their male peers to select the first story, *Let Me Promise You*, a story about the broken relationship between a young man and woman. The sixth story, *Murder on St. Valentine's Day* (a murder mystery) was also selected by a larger percentage of female than male twelfth graders. Conversely, male twelfth graders were more likely than their female counterparts to select story 3, *The Sniper*, a story about a young man caught in the perils of battle in the Irish civil war.

Approximately one-third of students at each grade who were allowed to choose a story indicated that they made their selections based on an anticipated emotional response or a general sense that the story would prove interesting or exciting. The second most frequently cited criteria for story selection was the title of the story or a review of the table of contents (which contained the authors' names as well as the stories' titles). Although selection criteria were quite similar at both grades 8 and 12, there was indication that twelfth graders were more likely to recognize the genre represented by a particular story and to make their selection based on that criteria.

In summary, patterns of students' choices in *The NAEP Reader* study indicated that students did select stories to read for the assessment that may be seen as representing to some extent their personal preferences. Several findings point to this conclusion. The popularity of certain stories was quite evident at each grade. There was little indication that most students simply read the first story or arbitrarily opened the booklet to any story: only 6 percent of eighth and twelfth graders indicated that the story's position in the book or the story's length or difficulty entered into their selection; and only 3 percent of eighth graders and 2 percent of twelfth graders indicated that no specific criterion guided their choice or that the selection was made randomly. Instead, students were likely to choose a story because they found something interesting about the story or expected to have a strong affective reaction to the story. The finding that twelfth graders

were more likely than eighth graders to base their story selection on genre may indicate a more sophisticated approach to reading for literary experiences among these students. This may be anticipated since, by the time students reach twelfth grade, they should have gained significantly more experiences with a broad range of literary genres than would be expected for students in the eighth grade.

Chapter 3

Context Effects on Constructed-Response Questions: Variations in Students' Responses Across Stories

To evaluate the impact of story selection in an assessment of reading comprehension *The NAEP Reader* study used comprehension questions that were generically worded so as to be applicable to any story chosen. The use of generic questions facilitated comparing students' comprehension by providing a standard set of questions that could be answered for each of the seven stories at both grades. Generically worded questions are often used by teachers in classrooms where students are given choices about what to read as a way of evaluating their comprehension of self-selected texts. For example, to prompt a student's retelling of a story, the teacher might ask "What was the main character's problem in the story?" Through such questioning, students' abilities and strategies can be assessed with materials they have chosen to read on their own.[15]

If the texts from which students can choose represent the same literary mode, specific text elements that may be used to frame comprehension questions are likely to be common across the texts. In fact, *The NAEP Reader* stories were, in part, selected to maximize the extent to which a common set of questions would be applicable to all of the stories. The seven stories at each grade were all fictional narratives involving characters, events, setting, conflict, and resolution. These common elements became the focus of comprehension questions that were worded without any reference to a specific story's content. For example, in order to assess students' understanding of character motivation, the following constructed-response question was worded so that it could be answered about any of the seven stories at each grade: "Describe what happens in this story that causes or motivates one of the main characters to act the way he or she does." (All of *The NAEP Reader* comprehension questions are presented in Appendix A.)

---

[15] Angeletti, S. R. (1991). Encouraging students to think about what they read. *The Reading Teacher, 45*(4), 288-296.

Filippo, R. F. (1997). *Reading assessment and instruction: A qualitative approach to diagnosis.* Harcourt Brace College Publishers: Fort Worth, TX.

While the stories in *The NAEP Reader* were chosen to accommodate generically worded questions, the stories were in fact far from identical. Therefore, one important factor influencing whether story choice can be used in large-scale assessments is the extent to which the context of the story changes the way a question functions. That is, generically worded questions may function as different questions in the context of certain stories than in the context of other stories. For example, they may be easier or more difficult, or they may rank students in a different manner depending on the story.

This chapter presents a close examination of students' performance on selected constructed-response questions across the seven stories at each grade. Two types of constructed-response questions were included in the assessment: short constructed-response questions that required no more than one or two sentences to answer, and extended constructed-response questions that typically required a paragraph. Of the eleven comprehension questions in *The NAEP Reader* study, eight were short constructed-response and three were extended constructed-response. Students' answers to short constructed-response questions were scored dichotomously, as either "unacceptable" or "acceptable." Answers to extended constructed-response questions were rated according to a four-level scoring rubric as "unsatisfactory," "partial," "essential," or "extensive." (Examples of scoring rubrics are presented in Appendix B.)

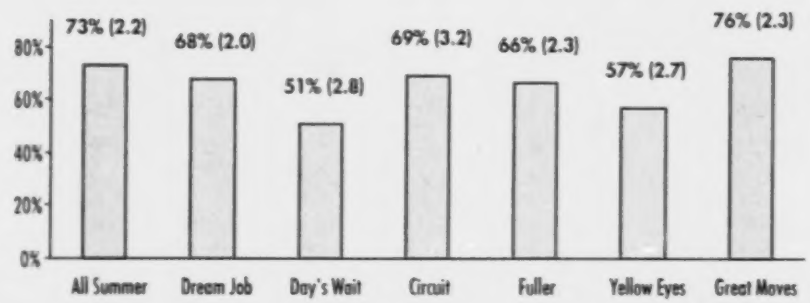
For the purpose of this examination, only the performance of students in the non-choice sample (i.e., those assigned a story to read) on selected comprehension questions is presented. Because the sample of students assigned to read each of the seven stories at both grades was nationally representative, the results for each story in the non-choice sample are representative of the population. Data from the choice sample, in which students were allowed to select a story, are not included in the results presented in this chapter since the choice sample for each individual story may not have been nationally representative and student performance on individual items may have been affected by their story selection.

The possibility that one or more stories were systematically harder or easier for all of the questions was explored; however, there was little evidence of this. Data for all of the questions by story for both grades are presented in Appendix C. The discussion presented in this chapter focuses on a sample of the questions for which variable patterns of student performance were observed across the seven stories at each grade.

Eighth Graders' Performance on Comprehension Questions Across Stories

On several of the eleven comprehension questions that accompanied each story, significant variation in difficulty was observed across the seven stories. Although the questions were worded identically, some either proved more difficult or easier in conjunction with one story than with another. Of all the short constructed-response questions, the one that most consistently had a relatively high percentage of acceptable responses across all of the grade eight stories was question #10: "Given what happened in this story, did you expect it to end the way it did? Give examples to support your answer." Across all the stories, slightly more than half to slightly more than three-quarters of students provided acceptable responses to this short constructed-response question. Although students had little difficulty with this question, there were still significant differences in how it functioned within the context of a story. As shown in Figure 3.1, a significantly higher percentage of students (76 percent) received a score of "acceptable" when answering this question about story 7, *Great Moves*, than when answering the same question about story 3, *A Day's Wait*, (51 percent).

Figure 3.1 **Percentages of Acceptable Responses to Short Constructed-Response Question #10 Across the Seven Stories at Grade 8**



The standard errors of the estimated percentages appear in parentheses. It can be said with 95 percent certainty that for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP) 1994 Reading Assessment.

Differences between the two stories may account for the different performance on this question. Story 7, *Great Moves*, is highly eventful, recounting the actions and interactions of two girlfriends with two boys who want a date and has a somewhat surprising ending when both girls refuse to go to the school dance because of the boys' behavior. The following response is typical of the 76 percent that were rated as "acceptable" for this story.

Sample Acceptable Response to Question #10 — *Great Moves*

Given what happened in this story, did you expect it to end the way it did? Give examples to support your answers.

No, I figured since both guys wanted both girls & both girls liked both guys, they would have just chosen & had a ~~dance~~ date for the dance.

In contrast to *Great Moves*, story 3 might be described as uneventful. The main action of *A Day's Wait* is circumscribed by the course of one day, the single setting of the boy's bedroom, and progresses mainly by sparse dialogue and description. Moreover, the major event of the story is a mental realization, not a physical action. Thus, a question beginning *Given what happened in the story...* may have presented some difficulty to students. Many student responses scored as "acceptable" relied on the fact of the temperature difference between Celsius and Fahrenheit to support their opinion. The following is a sample of the 51 percent of responses that were rated as "acceptable" for this story.

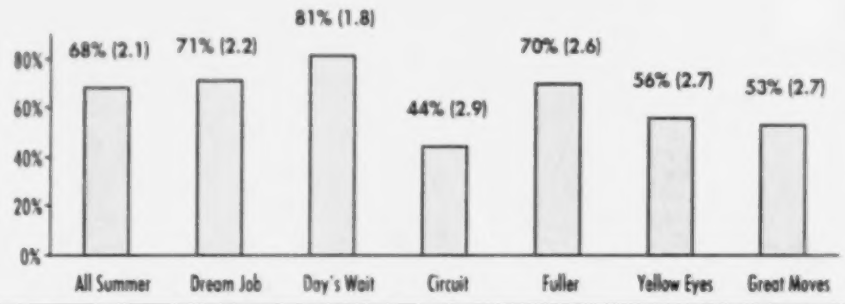
Sample Acceptable Response to Question #10 — *A Day's Wait*

Given what happened in this story, did you expect it to end the way it did? Give examples to support your answers.

yes, because I've heard that you usually don't die when you have influenza and a 102° temp.

Another question for which a large percentage of eighth-grade students provided acceptable responses on most stories was question #2, "Do you think this was a good title for this story? Tell why or why not." Across all the stories except one, more than half to more than three quarters of all students received a score of "acceptable" on this short constructed-response question. As shown in Figure 3.2, the highest percentage of students responded acceptably to this question in conjunction with story 3, *A Day's Wait.* Eighty-one percent of students assigned to read story 3 answered this question acceptably; this was significantly better performance than was shown by the students assigned to read any of the other stories in the eighth-grade *NAEP Reader.* The lowest percentage of students (44 percent) provided acceptable responses to this question in response to story 4, *The Circuit.*

Figure 3.2 **Percentages of Acceptable Responses to Short Constructed-Response Question #2 Across the Seven Stories at Grade 8**



The standard errors of the estimated percentages appear in parentheses. It can be said with 95 percent certainty that for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP) 1994 Reading Assessment.

A comparison of the two stories that elicited the highest and lowest percentage of acceptable responses may account for the wide margin in students' performance on this question. That students were most successful answering question #2 in conjunction with story 3, *A Day's Wait* might result from the fact that the title summarizes the subdued action of the story. Moreover, the narrator directly expresses "He had been waiting to die all day, ever since nine in the morning" on the story's last page. Responses scored as "acceptable" stated an opinion and supported it by critically evaluating the appropriateness of the title for the story. Responses scored as "unacceptable" stated an opinion with no meaningful support. The following is a sample response from among the 81 percent that were rated as "acceptable" for this story.

Sample Acceptable Responses to Question #2 — A Day's Wait

Do you think this was a good title for this story? Tell why or why not.

Yes, because Schatzy's waits a day to see if he is going to die or not. When he finds out he isn't going to die he relaxes.

A significantly lower percentage of students provided acceptable responses when evaluating the title of story 4, *The Circuit*. In contrast to the explicitness of *A Day's Wait*, as a title *The Circuit* bears a much more implicit, almost metaphorical connection to the story. The word "circuit" does not occur in the story, so to connect the continuous round of moving from job to job to the title required more interpretation. While most acceptable responses interpreted circuit in the sense of circular, some students incorporated prior knowledge in their evaluation of the title's appropriateness. Students were not required to interpret the metaphoric title in order to achieve an "acceptable" score. Responses supporting an opinion of why *The Circuit* was not a good title with information from the story about the little boy or about moving also were scored as "acceptable." The following response is an example of the 44 percent that were rated as "acceptable" for this story.

Sample Acceptable Response to Question #2 — *The Circuit*

Do you think this was a good title for this story? Tell why or why not.

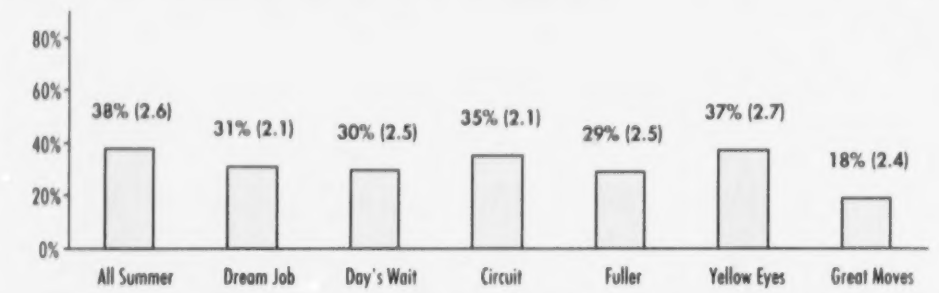
I think this title was good because it explained the moving and picking seasons. It's just like a curcuit, always going-round and round.

As mentioned earlier, three of the eleven questions that accompanied *The NAEP Reader* were extended constructed-response questions that required students to demonstrate their understanding by providing a more in-depth response. Of these three extended constructed-response questions, one was totally text-based and required students to explain a conflict in the story. The other two required students to take a personal stance toward a story by explaining its meaningfulness to them or by relating something from the story to their own experience. Performance of eighth graders was higher across all stories on the extended constructed-response question that required explaining a conflict in the story than on those that required a more personal interpretation.

Of the two extended constructed-response questions that required students to provide more personal responses, eighth graders were generally less successful answering question #11: "Tell how something in this story relates to something that happened to you." As shown in Figure 3.3, across all stories the percentages of students that achieved a score of "essential" or better ranged from 18 to 38 percent. The percentage of students receiving a score of "essential" or better on this question was highest for those reading story 1, *All Summer in a Day*, and lowest for those reading story 7, *Great Moves*.

Figure 3.3 **Percentages of Essential or Better Responses to Extended Constructed-Response Question #11 Across the Seven Stories at Grade 8**

THE NATION'S REPORT CARD



The standard errors of the estimated percentages appear in parentheses. It can be said with 95 percent certainty that for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP) 1994 Reading Assessment.

Surprisingly, students were better able to express a personal relation to a science fiction story that takes place on another planet than to a realistic story about having a date for the school dance. Responses scored "essential" or better provided an event from the story and an explanation of its personal relevance. The following is an example of the 38 percent of responses to *All Summer in A Day* that were rated as "essential" or better.

Sample Essential or Better Response to Question #11 — *All Summer in a Day*

Tell how something in this story relates to something that happened to you.

I can relate to Margot. One day it had been raining all day long so I could not go anywhere. Then the rain stopped but I missed it because I was doing something else. It then started again. Just like Margot had been trapped in the closet through the only sunshine in five years I missed the only sunshine that day.

The following is a sample response of the 18 percent that were rated as "essential" or better for the story, *Great Moves*.

Sample Essential or Better Response to Question #11 — *Great Moves*

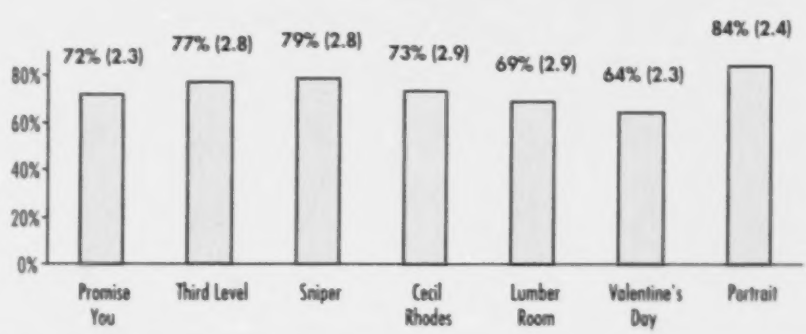
Tell how something in this story relates to something that happened to you.

I have had friends who were asked to a dance, yet still noticed and comforted my feelings of not having a date, the way Pooh treated Tigger. It was emotional and caring and it makes one like his/her friend much, much better. It helps you to trust them.

Twelfth Graders' Performance on Comprehension Questions Across Stories

Across all stories at grade 12, significant variation was seen in student performance on several of the eleven comprehension questions. Of the short constructed-response questions, twelfth graders demonstrated a fairly high rate of success across the seven stories with question #10, "Given what happened in this story, did you expect it to end the way it did? Give examples to support your answer." Across all stories, more than half to more than three-quarters of students received a score of "acceptable" on this short constructed-response question. As displayed in Figure 3.4, the highest percentage of students (84 percent) responded acceptably to this question in conjunction with story 7, *The Portrait*; and the lowest percentage of students (64 percent) provided acceptable responses in conjunction with story 6, *Murder on St. Valentine's Day*.

Figure 3.4 **Percentages of Acceptable Responses to Short Constructed-Response Question #10 Across the Seven Stories at Grade 12**



The standard errors of the estimated percentages appear in parentheses. It can be said with 95 percent certainty that for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP) 1994 Reading Assessment.

While this question presented little difficulty to students across all stories, the percentage of students' responses scored "acceptable" was significantly higher for story 7, *The Portrait*, than for story 6, *Murder on St. Valentine's Day*. All the action in *The Portrait* arises from the couple's being conned out of the only photograph of their son and proceeds according to Don Matteo's efforts to get the promised portrait. This linearity of plot might account for students' greater facility in making a connection between the beginning and the end. Acceptable responses supported their opinion with something that happened in the story. The following response to question #10 is a sample of the 84 percent that were rated as "acceptable" for the story, *The Portrait*.

Sample Acceptable Response to Question #10 — *The Portrait*

Given what happened in this story, did you expect it to end the way it did? Give examples to support your answers.

No, I was afraid that Mateo would never find the con artist and get retribution. It was also an unexpected twist when the artist was able to do the portrait from memory.

In comparison to story 7, the plot of *Murder on St. Valentine's Day* is more oblique. Although the action revolves around the face cream formula and a murder, implications of romance and characters' conversations decenter the reader's focus. As in any good mystery story, the reader is kept in suspense and needs to make more connections. The following is a sample of the 64 percent of responses that were rated as "acceptable" for this story.

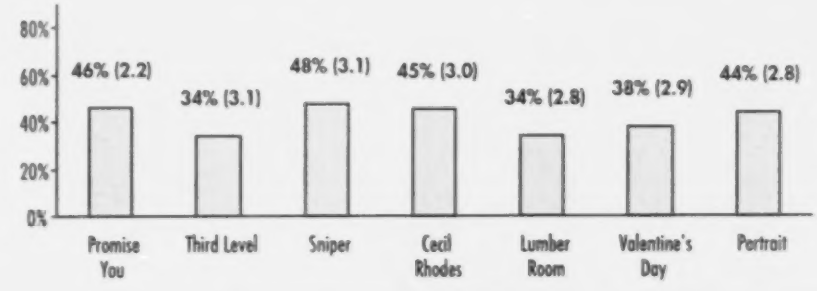
Sample Acceptable Response to Question #10 — *Murder on St. Valentine's Day*

Given what happened in this story, did you expect it to end the way it did? Give examples to support your answers.

Yes. When Miss Gray was living w/ Chrissa it was easier to figure out that she could be guilty because she didn't have a foolproof alibi. And since the formula couldn't be found the only unsearched thing was the box of stockings.

Of the short constructed-response questions, the one which appeared to be the most consistently difficult across the seven stories for twelfth-grade students was question #7: "Discuss the most outstanding qualities of one of the main characters in this story. Support your answer with examples from this story." Across all stories, less than half of twelfth graders provided an acceptable response. As shown in Figure 3.5, students assigned story 3, *The Sniper*, scored significantly higher on average with this question than those assigned story 2, *The Third Level*.

Figure 3.5 **Percentages of Acceptable Responses to Short Constructed-Response Question #7 Across the Seven Stories at Grade 12**



The standard errors of the estimated percentages appear in parentheses. It can be said with 95 percent certainty that for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP) 1994 Reading Assessment.

That students were more successful in discussing qualities about the main character in story 3, *The Sniper*, might be related to the context of war that, in itself, suggests certain character traits. Many responses mentioned either bravery or courage as one of the qualities. The following is a sample of the 48 percent of responses that were rated as "acceptable" for this story.

Sample Acceptable Response to Question #7 — *The Sniper*

Discuss the most outstanding qualities of one of the main characters in this story. Support your answer with examples from this story.

The most outstanding qualities of the republican sniper were his courage and his intelligence. He stood up to his enemy by returning his fire and then using his intelligence to outsmart the enemy.

In comparison, the situation in story 2, *The Third Level*, is rather unique and suggests no character traits inherent to the genre. Many responses mentioned only the character's determination, but did not infer from what he says that he must have a very active imagination. The following is a sample of the 34 percent of responses that were rated as "acceptable" for this story.

Sample Acceptable Response to Question #7 — *The Third Level*

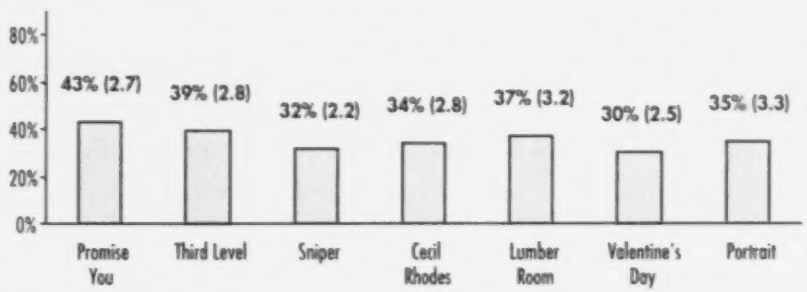
Discuss the most outstanding qualities of one of the main characters in this story. Support your answer with examples from this story.

Imagination and perserverence were Jerry's outstanding qualities. His imagination and perserverence kept him searching for "The Third Level"

Of the three extended constructed-response questions at grade 12, the one that proved to be most consistently difficult across the seven stories was question #11: "Tell how something in this story relates to something that happened to you." Across all stories, less than half of twelfth-graders achieved a score of "essential" or better. As displayed in Figure 3.6, a significantly higher percentage of students (43 percent) received a score of "essential" or better when responding about story 1 than the percentage of students (30 percent) responding to story 6.

Figure 3.6 **Percentages of Essential or Better Responses to Extended Constructed-Response Question #11 Across the Seven Stories at Grade 12**

THE NATION'S REPORT CARD



The standard errors of the estimated percentages appear in parentheses. It can be said with 95 percent certainty that for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP) 1994 Reading Assessment.

As the question requires making a relation to personal experience, students were more successful responding about a romance than about a murder mystery. The following sample response was among the 43 percent that were rated as "essential" or better for *Let Me Promise You.*

Sample Essential or Better Response to Question #11 — Let Me Promise You

Tell how something in this story relates to something that happened to you.

I relate to this story somewhat because at one time I, too felt that I deserved to be used, even if on a subconscious level. (When people believe they deserve such things, they try to go get them) I also could relate to her manipulative attitude because I have tried to get what I want from someone by using quiet tactics such as crying, anger and even material objects (gifts). This is why the character Alice really irritates and annoys me. Because she should have learned her lesson through this experience with Georgia, but instead "they lived happily ever after." The ending was unrealistic.

BEST COPY AVAILABLE

The following sample response was among the 30 percent that were rated as "essential" or better for the story, *Murder on St. Valentine's Day*.

Sample Essential or Better Response to Question #11 — *Murder on St. Valentine's Day*

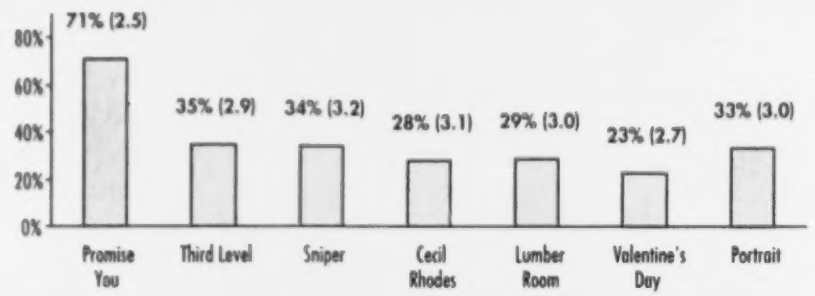
Tell how something in this story relates to something that happened to you.

I think the only possible thing that could relate to something that happened to me, is being stabbed in the back by a friend. When miss Carey tried to make it look like there was an affair happening between Clarissa and Mr. Marsh; she stabbed her friend in the back. Other than that I don't think there's much that could relate to something that happened to me.

BEST COPY AVAILABLE

One question at grade 12 functioned very differently in the context of one story as compared to all others. On short constructed-response question #12, "What additional information would have helped you to understand this story?", performance was low in conjunction with stories 2 through 7. As shown in Figure 3.7, the percentages of students receiving a score of "acceptable" for these stories ranged from 23 to 35 percent. In responding to this question about story 1, however, 71 percent of students achieved a score of "acceptable." As story 1, *Let Me Promise You*, recounts an evening visit between a woman and her former boyfriend, there is in a sense another story that led up to this evening. Most acceptable responses expressed that understanding would have been helped by knowing more about the couple's past relationship.

Figure 3.7 **Percentages of Acceptable Responses to Short Constructed-Response Question #12 Across The Seven Stories at Grade 12**



The standard errors of the estimated percentages appear in parentheses. It can be said with 95 percent certainty that for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP) 1994 Reading Assessment.

The following is a sample of the 71 percent of responses that were rated as "acceptable" for the story, *Let Me Promise You.*

Sample Acceptable Response to Question #12 — *Let Me Promise You*

What additional information would have helped you to understand this story?

additional information that would have helped me understand this story is what Alice and Georgis relationship was like in the past.

Summary and Discussion

Although textual elements common to all *The NAEP Reader* stories were considered in composing the comprehension questions, and although the same questions were used for students reading any of the stories, student performance varied significantly in the context of different stories. Despite the comparability of the stories in terms of grade appropriateness and overall difficulty, it was perhaps inevitable that textual elements could not be exactly equivalent across stories. The variety inherent in literary narratives presupposes differing emphases on character, setting, and action; titles of works can bear a literal or more metaphoric relation to stories; and stories can have more or less relation to a reader's experience. As the results presented in this chapter suggest, the difficulty of a question resides not only in the question itself, but also in the question's interaction with a particular text.

Comparing the performance of students who have read different stories is complicated by the fact of choice itself: students may not always select the stories they are best able to interpret. The volatility of identical questions when conjoined with different texts further complicates comparison of student performance in an assessment of reading comprehension. As evidenced in the data presented in this chapter, the same question may not elicit similar levels of comprehension performance when answered about different stories. Thus, allowing for the inevitable disparity between literary texts, it would appear that the challenges to making comparisons between students inherent when story selection is introduced in reading assessment may be further complicated by the variability of the text and question interaction.

Conclusion

The NAEP Reader study was an exploration of the effects of allowing students to select texts on an assessment of reading comprehension. In recognition of the importance of engagement, interest, and motivation in the processes of comprehension, self-selection of reading material has become a vital part of the reading curriculum. It has been suggested that allowing students to select texts not only increases their engagement in the reading process, but also promotes the development of reading interests that may translate into a life-long desire to pursue reading as a recreational activity and as a tool for gaining knowledge.

Concern for the authenticity and validity of reading assessments has led some educators to question the practice of measuring students' achievement in reading with texts that are assigned rather than self-selected. It has been suggested that if self-selection of reading material promotes interest and motivation then an assessment in which all students are assigned to read the same texts may not produce results that adequately reflect the extent of students' reading abilities. It is this concern that *The NAEP Reader* study was designed to address.

Summary of Findings

The results of this study do not provide evidence that eighth- and twelfth-grade students' performance on a reading assessment increases when they are given the opportunity to select a story from a collection of stories. In fact, there was some indication that the reading scores of eighth graders were on average slightly lower when they were allowed to select a story. While any generalizations made from these results must be constrained to the specific conditions of this study (e.g., the limited number of stories from which students could select and the use of only literary texts), it is clear that little or no support was provided by this study for the practice of text selection in a large-scale assessment of reading comprehension.

Students' reports about their perceptions of the assessment in the choice and the non-choice conditions shed some light on why performance was not higher for students who selected stories, but at the same time lead to further questions. Eighth and twelfth

graders in the choice sample were more likely than students in the non-choice sample to perceive the assessment as easier in comparison to other tests and assignments. Also, twelfth graders in the choice sample reported a higher estimation of their performance on the assessment than did their counterparts in the non-choice sample. While these findings may suggest that students felt more positive about the experience when they were allowed to select a story, there were no differences between choice and non-choice samples in students' reports of their motivation to perform well on the assessment.

It appears that allowing for choice may have somewhat altered students' impressions of the assessment, but did not result in increased efforts or, ultimately, in increased performance. Students in both the choice and the non-choice samples reported comparable levels of motivation, which may be one reason why few differences were observed in their scale scores. Why a more positive impression of the assessment, however, did not lead to increased motivation for performing well is one question that could not be addressed by this study. These findings should not be viewed as contradicting the research on motivation and comprehension. While increased levels of motivation may improve students' performance, there was no indication from these findings that allowing for text-selection heightened students' motivation in a large-scale assessment situation.

The patterns of story selections observed at both grades 8 and 12 suggest that students did take advantage of the opportunity to choose and, particularly at the twelfth grade, demonstrated definite preferences for certain types of stories. Although there was little indication that story selection patterns varied by racial/ethnic subgroups at either grade or by gender at grade 8, twelfth-grade males and females displayed very distinct preferences in the stories they chose to read. At grade 12, males were predominantly drawn to a story about a soldier and females were predominantly drawn to a story about a relationship. It is of interest that no significant differences between choice and non-choice performance was observed at grade 12, despite the overwhelming preferences demonstrated by twelfth-grade males and females for certain stories.

When students in the choice sample were asked to explain why they selected a particular story, the overwhelming response at both grades indicated an affective or general evaluative basis for their choice. Twelfth graders were more likely than eighth graders to base their selection on the topic or genre of the story, and they were more likely to indicate that they had read the story summaries included at the beginning of *The NAEP Reader* booklet or had skimmed the stories before making their decision. Eighth graders were more likely to report that it was a personal identification with something in the story, or some element of the story's characters, plot, or setting, that led to their choice.

To facilitate comparisons between students' understanding of different stories, *The NAEP Reader* study included eleven comprehension questions that were worded generically and could be answered about any one of the seven stories at each grade. The

results of this study, however, indicate that identical comprehension questions do not necessarily result in comparable levels of performance in the context of different stories. Although the stories that were included in the *The NAEP Reader* booklet at each grade were determined to be comparable in difficulty, variations in story elements made some questions more difficult or easier to answer. This finding points to a potential problem in allowing for story selection in large-scale assessments of reading comprehension. If identical questions function differently in the context of different stories, any comparison between students' performance with self-selected texts based on the same set of questions may not be valid.

Study Design Issues and Alternative Methods

The NAEP Reader study was designed to address issues related to choice on a large-scale standardized assessment of reading comprehension. As such, the procedures and materials conformed to the typical requirements of a large-scale testing program. The study was conducted concurrently with the administration of the main NAEP reading assessment; that is, sampled students were administered *The NAEP Reader* study within groups of students who were participating in the main assessment. Consequently, students taking *The NAEP Reader* were subject to the same timing limits and administrative procedures as students taking the main assessment. Conducting the study in this manner was consistent with the primary goal of the study to examine the effects of choice in a large-scale assessment situation.

As with any research study, it is important to understand how the design of the study may limit interpretations of the findings. Although this study was designed to answer specific questions about the effects of choice on assessment performance, it was not designed to address broader questions of how self-selection, motivation, and comprehension may be interrelated. Therefore, it would not be appropriate to consider the findings of this study as evidence that choice is unrelated to performance in other contexts, for other purposes, or with other tasks. Instead, this study should be seen as providing one piece of evidence that allowing students to make story selections may have little effect on their scores in a large-scale standardized assessment of reading comprehension.

Interpreting these findings should take into account the specific aspects of the study that preclude generalizing the results to different types of assessment procedures or materials. For example, the choice task in this study was limited to seven stories that were contained in *The NAEP Reader* booklets for each grade. Consequently, it may not have provided all students in the choice sample with a selection that most represented their personal interests and preferences. Thus, the choice task in *The NAEP Reader* should not

be viewed as comparable to other situations in which students have a more extensive range of texts and text types from which to make reading selections.

Another design factor that should be considered in interpreting the results of this study is the amount of time given to students in both the choice and non-choice samples. Both groups of students were given the same 50-minute time period that was also given to students in the main reading assessment. One potential concern regarding the standardized timing of both the choice and non-choice tasks is whether or not students in the choice sample were disadvantaged because they had the additional task of selecting a story and explaining the reason for their selection (a short constructed-response question) during the same 50-minute time period. Although there was no evidence that students who selected a story were less likely to complete all of the questions than were students who were assigned a story, it is not fully known how the time spent in selecting a story and explaining the reason for their selection affected their responses. It is possible that the time spent in selecting a story took away from time spent in composing their answer to each comprehension question. In consequence, students conscientiously attempting to provide a response to all the questions may have suffered a disadvantage and, though completing all the questions, provided less-developed answers by doing so.

Another specific aspect of this study's design that should be considered in interpreting the results is the use of generically worded questions to assess students' comprehension on each of the seven stories that were either assigned or selected. This study's use of generic questions, worded so as to be applicable to any of the seven stories at each grade, was meant to provide a standard basis for comparing students' comprehension across different stories. However, as described in Chapter 3 of this report, many of the generic questions demonstrated variable response patterns across the seven stories. That is, some questions were significantly more or less difficult to answer in the context of certain stories. As discussed in that chapter, the reason for this variability could be at least partially attributed to differing emphases in the interplay of narrative elements in each story.

Although this study provided information about the effects of choice under certain conditions, alternative designs could be considered for future studies that would provide additional information about how choice affects performance under other conditions. Although alternative study designs may present other types of limitations on the interpretations of findings, it may be useful to consider possible different approaches for future studies.

One alternative design would be to adjust the timing constraints to ensure that students in the choice samples are given time comparable to students in the non-choice sample for reading the story and answering the comprehension questions. Various methods could be used to achieve this. For instance, the study could be conducted without any timing constraints. Obviously, the absence of timing constraints would allow students

to select and read a story and to answer comprehension questions at their own pace. This procedure, however, would result in findings that have little applicability to standardized testing situations where timing constraints are necessary to administrative procedure. Another possibility would be to increase the time limits for both choice and non-choice samples to an extent that would ensure nearly every student had ample time to select, explain their selection, and complete the assessment. With this procedure, however, the uncertainty of how much time students in the choice sample use to make and explain the reason for their selections in relation to the amount of time they devote to reading the story and answering the comprehension questions would remain an issue. That is, the question of whether or not students' in the non-choice sample use more time than students in the choice sample for responding to each question would remain unanswered.

Yet another possible design would be to give students in the choice sample an extra 5 to 10 minutes before the assessment begins to make their selections and to give an explanation. The time allowed for reading the selected story and answering the comprehension questions would then be equivalent for both the choice and non-choice samples. Although this design would ensure that the processes of selecting a story and explaining the reason for that choice do not detract from the time spent in reading and answering questions, it is possible that other threats to validity may be introduced by this design. If students in the choice sample are given extra time to complete the selection process before the assessment, this may provide them with some advantages not afforded students in the non-choice sample. For example, students who spend time thinking about which story to read may have more of an opportunity to orient themselves to the assessment task and to the selected text. These students may begin to construct a preliminary mental model of the text to be read that might facilitate their comprehension once they begin reading and answering questions.

In addition to alternative timing procedures, future studies may attempt to control for the question/text interaction that was apparent with several of the generically worded questions used in this study. It may be possible to avoid the variability of question/text interactions by adjusting the demands of certain comprehension questions in light of specific text attributes, thus, increasing the comparability of questions across stories. However, such an effort would require extensive investigation of the comprehension processes elicited by different versions of the same question in the context of different stories to ensure their comparability. Furthermore, simply establishing the cognitive equivalence of different versions of questions *a priori* may not guarantee that individual students would respond in a predictable manner. Another possible method for ensuring that individual questions function comparably across stories would be to use fewer stories that are more similar to each other. Although this would probably decrease the variability of question/text interactions across stories, it would result in a severely

limited range of texts that are even less likely to reflect a range of students' interests and preferences.

One other design alternative that may be considered for future studies would be to first administer a common measure of reading ability across all sampled students who would then participate in *The NAEP Reader* assessment. In the design used for *The NAEP Reader* study, the basis for comparing students' performance across stories and between samples was the assumption that the groups of students being compared were each representative samples of eighth or twelfth graders in the nation. A standard measure of reading ability administered to all the students would provide direct substantiation of this assumption. Such a measure could provide information about the difficulty of passages and questions, and it would facilitate analyses of the interaction between reading ability and selection patterns.

Despite the limitations of this study as designed and administered, it is clear that the findings do not provide evidence that introducing text selection on an assessment of reading comprehension results in higher levels of performance. The implications for large-scale assessments of reading comprehension should be considered. Making comparisons between students is typically a primary objective of standardized assessments; however, introducing story selection as a part of the procedure results in comparisons that are not based on students' comprehension of the same reading material. Thus, differences in students' scores may be, in part, due to differences in the texts that were chosen. When allowing for story selection on an assessment of reading comprehension, analysis procedures become more complex and the comparisons made between students' performance based on their understanding of different stories become more tenuous. Given that the validity of test results are jeopardized when assessment procedures allow for choice, the results of this study suggest that what is lost by allowing for choice on a standardized assessment of reading comprehension may outweigh any perceived benefits.

Appendix A

The NAEP Reader Comprehension Questions

This appendix presents the reading comprehension questions that were administered in *The NAEP Reader* study. All students participating in the study were given questions two through eleven. Those students who were given a choice as to which story to read were asked to respond to an additional question (question #1) that asked them to indicate what story they had chosen and to explain why they chose the story.

*Grade 8**

1. Which story did you choose to read?

- Ⓐ All Summer in a Day
- Ⓑ Dream Job
- Ⓒ A Day's Wait
- Ⓓ The Circuit
- Ⓔ The Fuller Brush Man
- Ⓕ The Boy with Yellow Eyes
- Ⓖ Great Moves

Explain why you chose this story.

_____

_____

_____

_____

_____

*Grade 12**

1. Which story did you choose to read?

- Ⓐ Let Me Promise You
- Ⓑ The Third Level
- Ⓒ The Sniper
- Ⓓ Cecil Rhodes and the Shark
- Ⓔ The Lumber Room
- Ⓕ Murder on St. Valentine's Day
- Ⓖ The Portrait

Explain why you chose this story.

_____

_____

_____

_____

_____

*Question #1 was admnistered only to the sample of students who were given a choice of which story to read.

Grades 8 and 12

2. Do you think this was a good title for this story? Tell why or why not.

3. What do you find most meaningful about this story? Explain why.

4. What feeling was created by the language, setting, and major events in this story? Give specific examples.

5. Could the events have taken place in another setting? Tell why or why not.

6. Choose a conflict in this story and explain what the conflict is about.

7. Discuss the most outstanding qualities of one of the main characters in this story. Support your answer with examples from this story.

_____

_____

_____

_____

_____

8. Describe what happens in this story that causes or motivates one of the main characters to act the way he or she does.

_____

_____

_____

_____

_____

9. Choose an important event in the story and explain why it is crucial to this story's ending.

_____

_____

_____

_____

_____

10. Given what happened in this story, did you expect it to end the way it did? Give examples to support your answer.

_____

_____

_____

_____

_____

11. Tell how something in this story relates to something that happened to you.

12. What additional information would have helped you to understand this story?

Appendix B

Procedures and Methods

The 1994 *NAEP Reader* study was conducted to augment the 1994 NAEP reading assessment. The 1994 assessment continued the innovations that originated from the framework developed under the direction of the National Assessment Governing Board for the 1992 NAEP reading assessment. The framework views reading as a dynamic and interactive process. The assessment relied heavily on constructed-response questions to assess students' ability to interpret, to respond personally to, and to think critically about the text. Three purposes for reading were measured: reading for literary experience, reading to gain information, and reading to perform a task. *The NAEP Reader* study relied totally on constructed-response questions and assessed students' ability to read for literary experience under two different conditions: that of being able to select a story and that of being assigned a story to read.

Developing The NAEP Reader Study

The NAEP Reader study extended innovations in large-scale assessment by drawing on current research and practices in classroom literacy assessment to construct an instrument implementing those concepts at a national level. The Reading Instrument Development Committee for the 1992 NAEP assessment oversaw the development of both the field test and the operational versions of the special study. The development committee was comprised of leading researchers and educators in the field of reading. The members of the development committee are presented in Figure B.1. The committee held a total of four meetings to oversee the development, implementation, analysis, and reporting of the 1994 reading assessment and *The NAEP Reader* study.

Figure B.1 **The Reading Instrument Development Committee**

Dr. Mary Barr	Dr. Judith Langer
Dr. Carita Chapman	Edye Norniella
Dr. Richard Halle	Dr. Charles Peters
Dr. Elfrieda Heibert	Dr. John Pikulski
Dr. Barbara Kapinus	Dr. Robert Swartz

A field test of *The NAEP Reader* study was conducted in 1991 with a sample of approximately 500 students. The field test version of *The NAEP Reader* consisted of six stories at each grade and 14 comprehension questions. The stories had been selected from a pool of literary texts that had been submitted by reading educators across the country as representative of the type of stories used in typical eighth- and twelfth-grade classrooms. The Reading Instrument Development Committee was responsible for the final selection of stories included in the special study. Based on the results of that field test and information from field administrators, the committee made revisions to improve the instrument; two comprehension questions were deleted and several questions were revised. Also, an additional story was added to *The NAEP Reader* at both grades eight and twelve, expanding the collection's cultural diversity.

A pilot test of The NAEP Reader study was administered in 1992 with nationally representative samples of 2,138 eighth graders and 1,918 twelfth graders. All students involved in the pilot test were allowed to select a story. However, no comparisons between choice and non-choice performance could be made based on the results of this pilot because there had not been an attempt to administer the NAEP Reader stories and comprehension questions under a non-choice condition.

Design of The 1994 NAEP Reader Study

The NAEP Reader study was developed to gather information about the effects of choice on student performance in an assessment of reading comprehension. *The NAEP Reader* was administered as a part of the main 1994 NAEP reading assessment. Students participating in the study were assessed along with students who were sampled for the main assessment. Students in the study were given a collection of grade-appropriate stories reproduced in a booklet entitled *The NAEP Reader.* At each grade assessed, the collection included seven stories. In addition to the stories, the booklet contained a table of contents listing the seven stories and their authors, as well as summaries that appeared on the inside front cover and provided students with some idea of the characters and plot in each story. (The story summaries are reproduced in the Introduction of this report.) In addition, all students received separate booklets containing background questions, questions about their motivation, and constructed-response comprehension questions (eight short constructed-response and three extended constructed-response questions).

In order to examine the effects of choice, a nationally representative sample of eighth and twelfth graders were allowed to select which story to read. In addition, eighth- and twelfth-grade students in nationally representative samples were assigned one of the stories to read (one sample for each story at each grade). This design allowed for the comparison of the choice and non-choice samples based on the assumption that the overall ability of the two parallel populations only differs due to measurable sampling variability. The inclusion of a non-choice sample for each of the seven stories at each grade makes it possible to compare the performance of students who had selected to read a story (a subset of a representative sample) with that of a nationally representative sample of students (a parallel population) who were assigned to read the same story. The performance of students in the non-choice samples served as the standard against which choice performance was compared.

Both samples of students (choice and non-choice) received the same booklet of stories and the same background, motivation, and comprehension questions. There was one exception: students in the choice sample were asked to explain why they had chosen a particular story. Students in the non-choice sample were told by the test administrator which story to read. Students in the choice sample were instructed to select a story, and were given an additional question asking them to indicate the story they selected and to briefly explain why they chose it. The assessment time in both the choice and non-choice conditions was 50 minutes.

Sampling

The results presented in this report are based on nationally representative probability samples of eighth- and twelfth-grade students. The samples were selected using a complex multistage sampling design involving the sampling of students from selected schools within selected geographic areas across the country. The sample design had the following stages;

1. selection of primary sampling units which were geographic areas (counties or groups of counties);
2. selection of schools (both public and nonpublic) within the selected areas: and
3. selection of students within selected schools.

Each selected school that participated in the assessment, and each student assessed, represents a portion of the population of interest. To make valid inferences from the student samples to the respective populations from which they were drawn, sampling weights are needed. Sampling weights account for disproportionate representation due to oversampling of students attending schools with a high concentration of Black and/or Hispanic students, and from nonpublic schools. Lower sampling rates for very small schools must also be accounted for with the weights.

Table B.1 provides a summary of the weighted student sample sizes for each story in *The NAEP Reader* study in both the choice and non-choice samples. The numbers reported include both public and nonpublic school students.

Table B.1 **Weighted Choice and Non-Choice Sample Sizes by Grade and Story**

	Total	Story 1	Story 2	Story 3	Story 4	Story 5	Story 6	Story 7
8th Grade								
Choice	2,416 (100%)	661 (27.4%)	452 (18.7%)	424 (17.6%)	106 (4.4%)	78 (3.2%)	441 (18.3%)	254 (10.5%)
Non-Choice	4,825 (100%)	859 (17.8%)	731 (15.2%)	741 (15.4%)	581 (12.0%)	592 (12.3%)	667 (13.8%)	654 (13.6%)
12th Grade								
Choice	2,100 (100%)	533 (25.4%)	166 (7.9%)	576 (27.4%)	147 (7.0%)	66 (3.2%)	484 (23.0%)	128 (6.1%)
Non-Choice	3,664 (100%)	629 (17.2%)	536 (14.6%)	564 (15.4%)	456 (12.4%)	480 (13.1%)	537 (14.6%)	462 (12.6%)

The standard errors of the estimated percentages appear in parentheses. It can be said with 95 percent certainty that for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP) 1994 Reading Assessment.

Tables B.2 presents the demographic make-up of the nationally representative choice and non-choice samples for each of the seven stories at grade 8.

Table B.2

Sample Sizes and Weighted Percentages of Students by Race/Ethnicity and Gender in the Non-Choice Samples for Each Story and in the Total Non-Choice and Choice Samples: Grade 8

	Non-Choice							Non-Choice	Choice
	Story 1	Story 2	Story 3	Story 4	Story 5	Story 6	Story 7	Total	Total
White	559(65.0%)	489(66.9%)	506(68.2%)	396(68.2%)	422(71.3%)	467(70.0%)	470(71.9%)	3309(68.6%)	1610(66.7%)
Black	142(16.5%)	130(17.8%)	117(15.9%)	102(17.6%)	78(13.1%)	96(14.4%)	89(13.6%)	755(15.6%)	408(16.9%)
Hispanic	111(12.9%)	78(10.6%)	83(11.2%)	60(10.4%)	61(10.4%)	69(10.3%)	73(11.1%)	535(11.1%)	292(12.1%)
Asian	20(2.4%)	12(1.6%)	9(1.3%)	7(1.3%)	14(2.3%)	13(2.0%)	6(1.0%)	82(1.7%)	30(1.2%)
Pacific Islander	13(1.5%)	5(0.7%)	11(1.5%)	6(1.0%)	7(1.2%)	7(1.0%)	6(0.9%)	55(1.1%)	26(1.1%)
American Indian	13(1.5%)	13(1.7%)	13(1.8%)	7(1.2%)	10(1.6%)	12(1.8%)	8(1.2%)	75(1.6%)	44(1.8%)
Male	447(52.1%)	379(51.8%)	405(54.6%)	330(56.9%)	313(52.9%)	343(51.4%)	348(53.3%)	2565(53.2%)	1288(53.3%)
Female	412(47.9%)	352(48.2%)	336(45.4%)	250(43.1)	279(47.1%)	324(48.6%)	306(46.7%)	2260(46.8%)	1128(46.7%)

The standard errors of the estimated percentages appear in parentheses. It can be said with 95 percent certainty that for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP) 1994 Reading Assessment

BEST COPY AVAILABLE

Tables B.3 presents the demographic make-up of the nationally representative choice and non-choice samples for each of the seven stories at grade 12.

Table B.3	Sample Sizes and Weighted Percentages of Students by Race/Ethnicity and Gender in the Non-Choice Samples for Each Story and in the Total Non-Choice and Choice Samples: Grade 12	

	Non-Choice							Non-Choice	Choice
	Story 1	Story 2	Story 3	Story 4	Story 5	Story 6	Story 7	Total	Total
White	427 (67.8%)	384 (71.7%)	405 (71.7%)	335 (73.4%)	340 (70.8%)	396 (73.7%)	336 (72.8%)	2622 (71.6%)	1470 (70.0%)
Black	105 (16.6%)	79 (14.7%)	77 (13.7%)	69 (15.1%)	68 (14.1%)	72 (13.4%)	62 (13.4%)	531 (14.5%)	318 (15.1%)
Hispanic	69 (10.9%)	51 (9.6%)	54 (9.6%)	41 (9.0%)	48 (10.0%)	51 (9.4%)	37 (8.0%)	351 (7.6%)	208 (9.9%)
Asian	18 (2.9%)	14 (2.6%)	11 (2.0%)	7 (1.6%)	10 (2.1%)	13 (2.5%)	16 (3.4%)	90 (2.5%)	57 (2.7%)
Pacific Islander	3 (0.4%)	3 (0.5%)	8 (1.5%)	2 (0.4%)	10 (2.1%)	1 (0.3%)	3 (0.6%)	30 (0.8%)	19 (0.9%)
American Indian	6 (0.9%)	4 (0.7%)	8 (1.3%)	3 (0.6%)	5 (0.9%)	3 (0.6%)	1 (0.3%)	29 (0.8%)	19 (0.9%)
Male	296 (47.1%)	292 (54.5%)	337 (59.6%)	256 (56.2%)	238 (49.5%)	282 (52.6%)	245 (53.1%)	1946 (53.1%)	1100 (52.4%)
Female	333 (52.9%)	244 (45.5%)	228 (40.4%)	200 (43.8%)	243 (50.5%)	255 (47.4%)	217 (46.9%)	1718 (46.9%)	1000 (47.6%)

The standard errors of the estimated percentages appear in parentheses. It can be said with 95 percent certainty that for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP) 1994 Reading Assessment

BEST COPY AVAILABLE

NAEP Reader Reporting Groups

This report contains results for the nation and for groups of students defined by shared characteristics. Because of the limited nature of *The NAEP Reader*'s study sample, not all of the traditional NAEP reporting groups are presented. The reporting subgroups presented in this report are race/ethnicity and gender and are defined as follows.

Race/Ethnicity

Results are presented for students in different racial/ethnic groups based on the students' self-identification of their race/ethnicity according to the following mutually exclusive categories: White, Black, Hispanic, Asian, Pacific Islander, and American Indian (which includes Alaskan Native). Data are reported for subpopulations only where sufficient numbers of students are present. For *The NAEP Reader* study, results are reported only for White, Black, and Hispanic subpopulations.

Gender

Results are reported separately for males and females.

Scoring

Materials from the 1994 *NAEP Reader* study were shipped to National Computer Systems in Iowa City, Iowa for processing. Receipt and quality control were managed through a sophisticated bar-coding and tracking system. After all appropriate materials were received from a school, they were forwarded to the professional scoring area where the student responses to the constructed-response questions were evaluated by trained staff using guidelines prepared by Educational Testing Service (ETS) staff. Each constructed-response question had a unique scoring guide that defined the criteria to be used in evaluating students' responses. In developing the scoring guides, a generic rubric was first written that served as the framework for creating story-specific versions of the rubrics for each story. The short response questions were scored as either acceptable or unacceptable. The extended constructed-response questions were evaluated with four-level scoring rubrics that permit partial credit to be given. The following two scoring guides are examples of the generic rubrics that were developed for creating the story-specific guides. The first is a short constructed-response guide and the second is an extended constructed-response guide.

Sample Short Constructed-Response Question Scoring Guide

Item 2. Do you think this was a good title for this story? Tell why or why not.

4 = Acceptable	Acceptable responses state an opinion. Students must understand the title of the story, and critically evaluate the appropriateness of the title of the story.
1 = Unacceptable	Unacceptable responses are incomplete. They either state an unsupported opinion, or state support for an opinion that cannot readily be inferred.

Sample Extended Constructed-Response Question Scoring Guide

Item 6. Choose a conflict in this story and explain what the conflict is about.

Stance:	Critical Stance
General Scoring Rubric:	Demonstrate understanding of a conflict in the story.
Scoring Rationale:	The task requires students to • provide evidence that they understand what a conflict is; • discuss an important conflict in the story.
1=Unsatisfactory	These responses provide inappropriate or inaccurate information.
2=Partial	These responses provide a conflict from the story without discussing it.
3=Essential	These responses provide a conflict and discuss it or provide a discussion for which the conflict is implied.
4=Extensive	These responses provide an important conflict and explain that conflict. These responses are much more interpretive in nature and move beyond a literal interpretation of the conflict.

As the same questions were administered across the seven stories at each grade level, the scoring criteria were applied within the context of each story. Training of scorers began with a general explanation of the scoring guide for each question and progressed sequentially applying the scoring criteria to each specific story. Thus, scorers were trained seven times on each unique scoring guide at each grade as the scoring sessions proceeded from one story to the next.

For *The NAEP Reader* study, approximately 180,000 student responses were scored. This figure includes a 25 percent rescore to monitor interrater reliability. The overall interrater percentages of agreement between scorers on each comprehension question for the 1994 national reliability samples ranged from 77 percent to 92 percent at grade 8, and from 81 percent to 92 percent at grade 12. The percentage agreement between scorers for the question asking students to describe why they selected a story was 91 percent at grade 8 and 93 percent at grade 12. Table B.4 presents the percentages of interrater exact agreement for each comprehension question across the seven stories administered to the non-choice samples at grades 8 and 12.

Table B.4 **Percentages of Exact Agreement Between First and Second Scorers on the Reliability Sample of Responses to Comprehension Questions**

Comprehension Question #		Grade 8	Grade 12
2	*(Short)*	92%	92%
3	*(Extended)*	77%	81%
4	*(Short)*	86%	88%
5	*(Short)*	87%	91%
6	*(Extended)*	81%	84%
7	*(Short)*	90%	89%
8	*(Short)*	87%	90%
9	*(Short)*	84%	86%
10	*(Short)*	87%	90%
11	*(Extended)*	90%	89%
12	*(Short)*	90%	91%

Data Analysis and IRT Scoring

Subsequent to the professional scoring, all constructed-response scoring and background questionnaire information was transcribed to the NAEP database at ETS. Each processing activity was conducted with rigorous quality control. After the assessment information had been compiled in the database, the data were weighted according to the population structure. The weighting probability of selection for each student as a result of the sampling design was adjusted for nonresponse. Through post-stratification, the weighting assured that the representation of certain subpopulations corresponded to figures from the U.S. Census and the Current Population Survey.[1]

Analyses were then conducted to determine the percentages of students who gave various responses to each cognitive and background question. In determining these percentages for the cognitive questions, a distinction was made between missing responses at the end of a block (i.e., missing responses subsequent to the last question the student answered) and missing responses prior to the last observed response. Missing responses before the last observed response were considered intentional omissions. Missing responses at the end of the block were considered "not reached" and treated as if the questions had not been presented to the student. In calculating response percentages for each question, only students classified as having been presented the question were included in the denominator of the statistic.

Item response theory (IRT) was used to estimate average reading scale scores for each story for the nation and for various subgroups of interest within the nation. IRT models the probability of answering a question in a certain way as a mathematical function of proficiency or skill. The main purpose of IRT analysis is to provide a common scale on which performance can be compared across groups such as those defined by grades and characteristics, including race/ethnicity and gender.

For the 1994 *NAEP Reader* study, a separate scale ranging from 0 to 100 was created to report performance for each story at each grade. The scale was first established based on the performance of students who were assigned a story in the non-choice samples. The scale parameters were then applied to students' performance on each story in the choice samples. The scale summarizes student performance across the two question types in the special study — short constructed-response and extended constructed-response. The scale has a mean of 50 and a standard deviation of 10.

---

[1] For additional information about the use of weighting procedures in NAEP, see Johnson, E. G. (1989, December). Considerations and techniques for the analysis of NAEP data. *Journal of Educational Statistics, 14*(4), 303-334.

In order to compare students' performance across stories, the scales were linked by making the means and standard deviations the same for each story. Because the stories had been randomly assigned across the non-choice sample, the subsample of students who received each story was representative of the population. Thus, it could be assumed that the distribution of students' scores would be equivalent across the seven stories within each grade.

In producing *The NAEP Reader* scale, two distinct IRT models were used. Short constructed-response questions rated as acceptable or unacceptable were scaled using the two-parameter logistic (2PL) model; and extended constructed-response questions, rated on a four-level rubric, were scaled using a generalized partial-credit (GPC) model. [2] Developed by ETS and first used in 1992, the GPC model permits the scaling of questions scored according to multipoint rating schemes. The model takes full advantage of the information available from each of the student response categories used for these more complex constructed-response questions.

Estimating Variability

Because the statistics presented in this report are estimates of group and subgroup performance based on samples of students, rather than the values that could be calculated if every student in the nation answered each question, it is important to account for the degree of uncertainty associated with the estimates. Two components of uncertainty are accounted for in the variability of statistics based on proficiency: 1) the uncertainty due to sampling only a relatively small number of students, and 2) the uncertainty due to sampling only a relatively small number of reading comprehension questions. The variability associated with the estimated percentages of students who answered a certain cognitive question correctly or responded in a certain way to a motivation question is accounted for by the first component alone.

In addition to providing estimates of percentages of students and their average scale scores, this report provides information about the uncertainty of each statistic. Because NAEP uses complex sampling procedures, conventional formulas for estimating sampling variability that assume simple random sampling are inappropriate. NAEP uses a jackknife replication procedure to estimate standard errors. The jackknife standard error provides a reasonable measure of uncertainty for any information about students that can be observed without error. However, each student responded to so few questions that the proficiency measurement for any single student is imprecise. In this case, using plausible values technology makes it possible to describe the performance of groups and

[2] Muraki, E. (1992). A generalized partial credit model: Application of an EM algorithm. *Applied Psychological Measurement, 16*(2), 159-176.

subgroups of students, but the underlying imprecisions that makes this step necessary adds an additional component of variability to statistics based on NAEP scale scores.[3]

The reader is reminded that, like findings from all surveys, *The NAEP Reader* results are also subject to other kinds of error, including the effects of imperfect adjustment for student and school nonresponse, and other unknowable effects associated with the particular instrumentation and data collection methods. Nonsampling errors can be attributed to a number of sources: inability to obtain complete information about all selected schools in the sample (some students or schools refused to participate, or students participated but answered only certain questions); ambiguous definitions; differences in interpreting questions; inability or unwillingness to give correct information; mistakes in recording, coding, or scoring data, and other errors of collecting, processing, sampling, and estimating missing data. The extent of nonsampling error is difficult to estimate. By their nature, the impact of such errors cannot be reflected in the data-based estimates of uncertainty provided in NAEP reports.

Drawing Inferences from the Results

The use of *confidence intervals*, based on the standard errors, provides a way to make inferences about the population means and percentages in a manner that reflects the uncertainty associated with the sample estimates.

An estimated sample mean scale score ± 2 standard errors represents a 95 percent confidence interval for the corresponding population quantity. This means that with approximately 95 percent certainty, the average performance of the entire population of interest is within ± 2 standard errors of the sample mean.

As an example, suppose that the average reading scale score of students in a particular group was 51, with a standard error of 1.2. A 95 percent confidence interval for the population quantity would be as follows:

Mean ± 2 standard errors

51 ± 2 x 1.2

51 ± 2.4

48.6, 53.4

Thus, one can conclude with 95 percent certainty that the average scale score for the entire population of students in that group is between 48.6 and 53.4.

---

[3] For further details, see Johnson, E. G., & Rust, K. F. (1992). Population inferences and variance estimation for NAEP data. *Journal of Educational Statistics, 17*(2), 175-190.

Similar confidence intervals can be constructed for percentages, provided that the percentages are not extremely large (greater than 90) or extremely small (less than 10). For extreme percentages, confidence intervals constructed in the above manner may not be appropriate. However, procedures for obtaining accurate confidence intervals are quite complicated. Thus, comparisons involving extreme percentages should be interpreted with this in mind.

To determine whether there is a real difference between the mean scale score (or percentage of a certain attribute) for two groups in the population, one needs to obtain an estimate of the degree of uncertainty associated with the difference between the scale score means or percentages of these groups for the sample. This estimate of the degree of uncertainty — called the standard error of the difference between the groups — is obtained by taking the square of each group's standard error, summing these squared standard errors, and then taking the square root of this sum.

$$\sqrt{SE_1^2 + SE_2^2}$$

Similar to the manner in which the standard error for an individual group mean or percentage is used, the standard error of the difference can be used to help determine whether differences between groups in the population are real. The difference between the mean scale score or percentage of the two groups $\pm$ 2 standard errors of the difference represents an approximate 95 percent confidence interval. If the resulting interval includes zero, there is insufficient evidence to claim a real difference between groups in the population. If the interval does not contain zero, the difference between groups is statistically significant (different) at the .05 level.

The procedures described in this section, and the certainty ascribed to intervals (e.g., 95 percent confidence intervals) are based on statistical theory that assumes that only one confidence interval or test of statistical significance is being performed. When one considers sets of confidence intervals, like those for the average scale scores for all racial/ethnic subgroups, statistical theory indicates that the certainty associated with the entire set of intervals is less than that attributable to each individual comparison from the set. If one wants to hold the certainty level for a specific set of comparisons at a particular level (e.g., 95), adjustments (called multiple-comparisons procedures) need to be made.

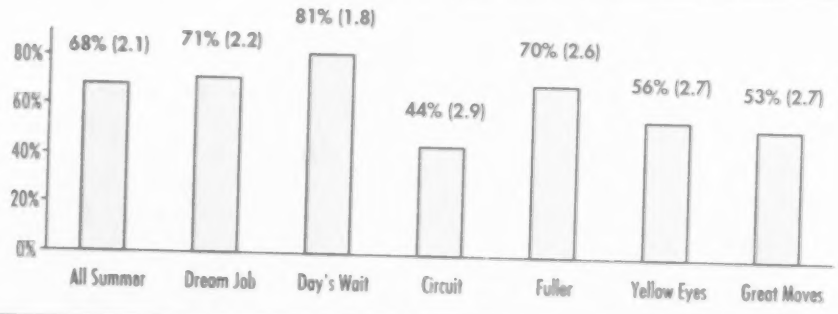
Appendix C

Student Performance on Comprehension Questions: Grade 8

This appendix presents figures that illustrate eighth-grade students' performance on the comprehension questions in *The NAEP Reader* study. For each of the 11 comprehension questions, the performance of students in the non-choice samples on each of the stories are presented. For short constructed-response questions (scored dichotomously as Acceptable or Unacceptable), the figures present the percentages of students who gave responses that were rated as Acceptable. For extended constructed-response questions (scored according to a four-level rubric as Unsatisfactory, Partial, Essential, or Extensive), the figures present the percentages of students who gave Essential or better responses.

Figure C.1 **Question #2: *Do you think this was a good title for this story? Tell why or why not. Grade 8***

THE NATION'S REPORT CARD



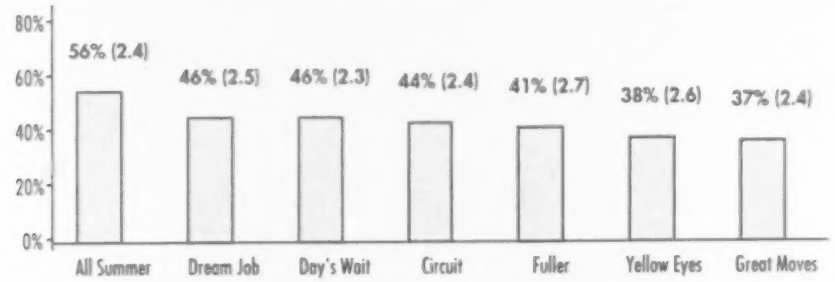
The standard errors of the estimated percentages appear in parentheses. It can be said with 95 percent certainty that for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP) 1994 Reading Assessment.

For short constructed-response questions, the data presented are percentages of students who gave Acceptable responses. For extended constructed-response questions, the data presented are percentages of students who gave Essential or better responses.

BEST COPY AVAILABLE

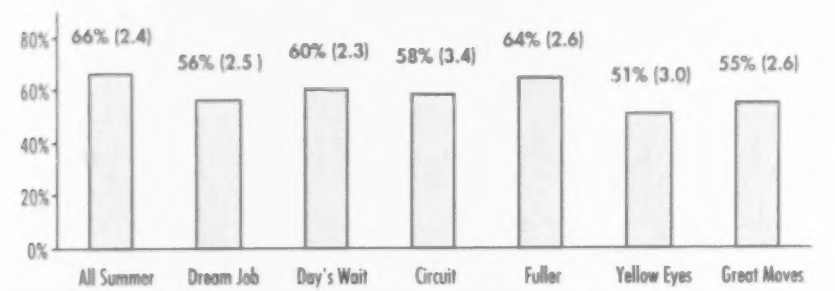
Figure C.2	**Question #3: *What do you find most meaningful about this story? Explain why.* Grade 8**	THE NATION'S REPORT CARD



The standard errors of the estimated percentages appear in parentheses. It can be said with 95 percent certainty that for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP) 1994 Reading Assessment

Figure C.3	**Question #4: *What feeling was created by the language, setting, and major events in this story? Give specific examples.* Grade 8**	THE NATION'S REPORT CARD

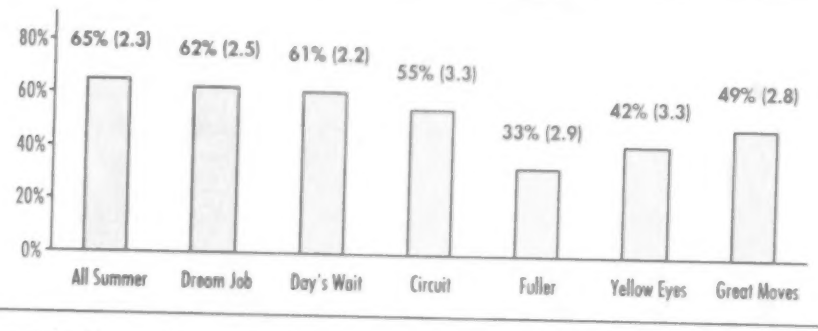


The standard errors of the estimated percentages appear in parentheses. It can be said with 95 percent certainty that for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP) 1994 Reading Assessment.

For short constructed-response questions, the data presented are percentages of students who gave Acceptable responses. For extended constructed-response questions, the data presented are percentages of students who gave Essential or better responses.

Figure C.4 **Question #5: *Could the events have taken place in another setting? Tell why or why not.* Grade 8**



The standard errors of the estimated percentages appear in parentheses. It can be said with 95 percent certainty that for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP) 1994 Reading Assessment.

Figure C.5 **Question #6: *Choose a conflict in this story and explain what the conflict is about.* Grade 8**

The standard errors of the estimated percentages appear in parentheses. It can be said with 95 percent certainty that for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP) 1994 Reading Assessment

---

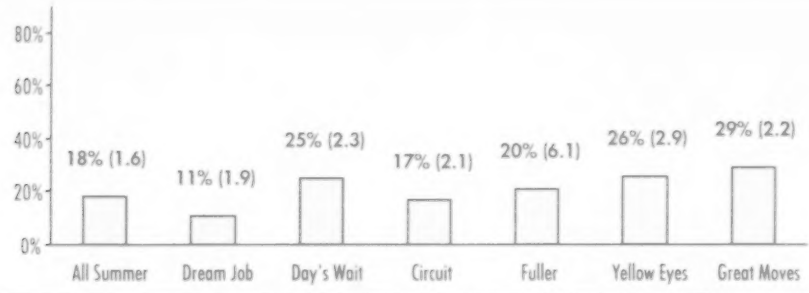
For short constructed-response questions, the data presented are percentages of students who gave Acceptable responses. For extended constructed-response questions, the data presented are percentages of students who gave Essential or better responses.

BEST COPY AVAILABLE

Figure C.6

Question #7: *Discuss the most outstanding qualities of one of the main characters in this story. Support your answer with examples from this story.* Grade 8

THE NATION'S REPORT CARD



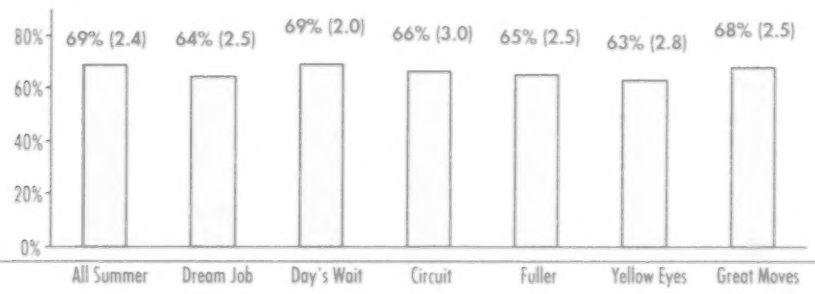
The standard errors of the estimated percentages appear in parentheses. It can be said with 95 percent certainty that for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP) 1994 Reading Assessment

Figure C.7

Question #8: *Describe what happens in this story that causes or motivates one of the main characters to act the way he or she does.* Grade 8

THE NATION'S REPORT CARD



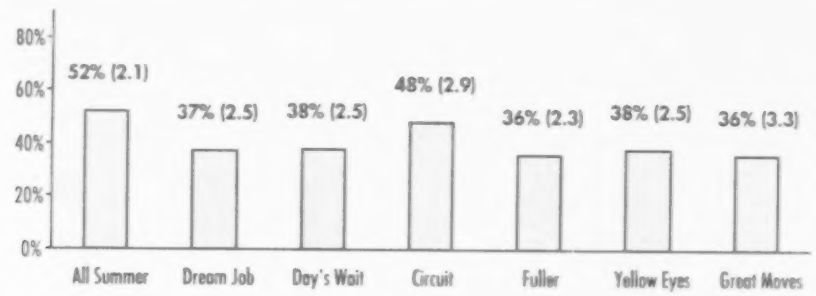
The standard errors of the estimated percentages appear in parentheses. It can be said with 95 percent certainty that for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP) 1994 Reading Assessment

---

For short constructed-response questions, the data presented are percentages of students who gave Acceptable responses. For extended constructed-response questions, the data presented are percentages of students who gave Essential or better responses.

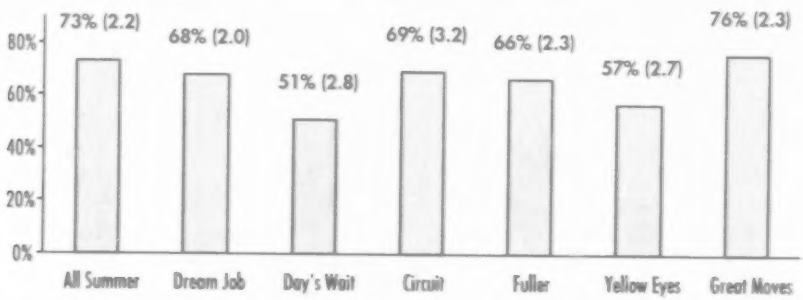
Figure C.8 **Question #9: *Choose an important event in the story and explain why it is crucial to this story's ending.* Grade 8**



The standard errors of the estimated percentages appear in parentheses. It can be said with 95 percent certainty that for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP) 1994 Reading Assessment

Figure C.9 **Question #10: *Given what happened in this story, did you expect it to end the way it did? Give examples to support your answers.* Grade 8**

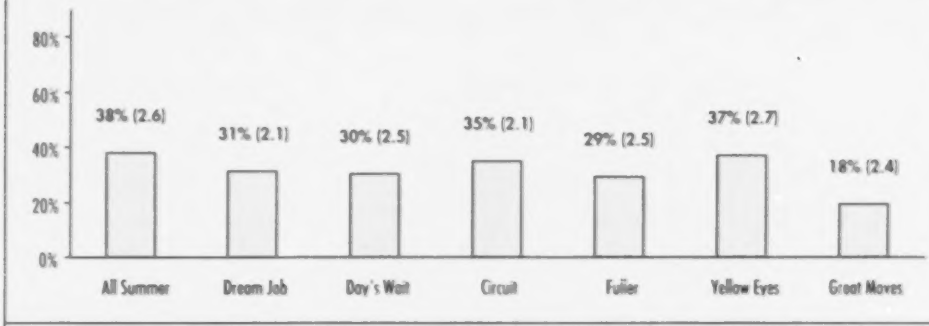


The standard errors of the estimated percentages appear in parentheses. It can be said with 95 percent certainty that for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP) 1994 Reading Assessment

For short constructed-response questions, the data presented are percentages of students who gave Acceptable responses. For extended constructed-response questions, the data presented are percentages of students who gave Essential or better responses.

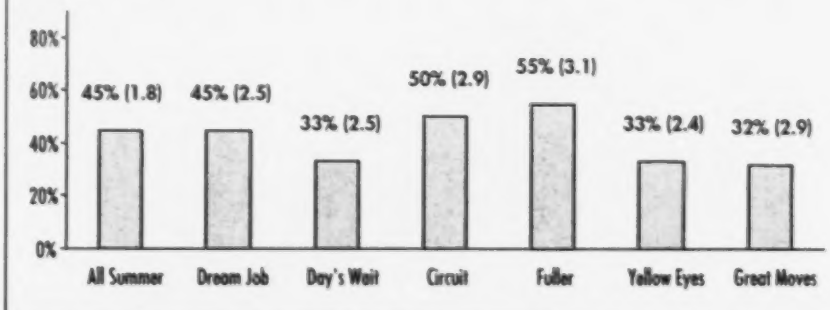
Figure C.10	**Question #11: *Tell how something in this story relates to something that happened to you.* Grade 8**	THE NATION'S REPORT CARD



The standard errors of the estimated percentages appear in parentheses. It can be said with 95 percent certainty that for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP) 1994 Reading Assessment

Figure C.11	**Question #12: *What additional information would have helped you to understand this story?* Grade 8**	THE NATION'S REPORT CARD



The standard errors of the estimated percentages appear in parentheses. It can be said with 95 percent certainty that for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP) 1994 Reading Assessment

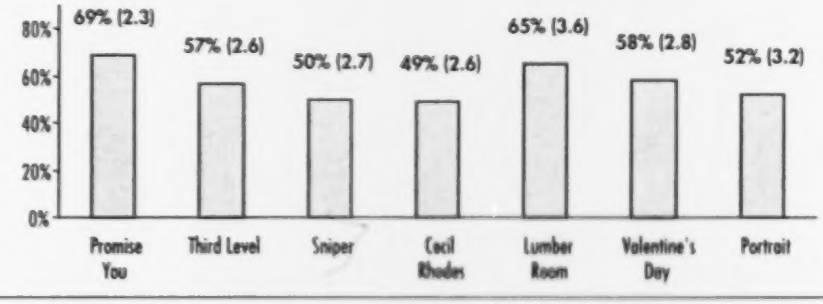
For short constructed-response questions, the data presented are percentages of students who gave Acceptable responses. For extended constructed-response questions, the data presented are percentages of students who gave Essential or better responses.

Appendix D

Student Performance on Comprehension Questions: Grade 12

This appendix presents figures that illustrate twelfth-grade students' performance on the comprehension questions in *The NAEP Reader* study. For each of the 11 comprehension questions, the performance of students in the non-choice samples on each of the stories are presented. For short constructed-response questions (scored dichotomously as Acceptable or Unacceptable), the figures present the percentages of students who gave responses that were rated as Acceptable. For extended constructed-response questions (scored according to a four-level rubric as Unsatisfactory, Partial, Essential, or Extensive), the figures present the percentages of students who gave Essential or better responses.

Figure D.1 **Question #2: *Do you think this was a good title for this story? Tell why or why not .* Grade 12**

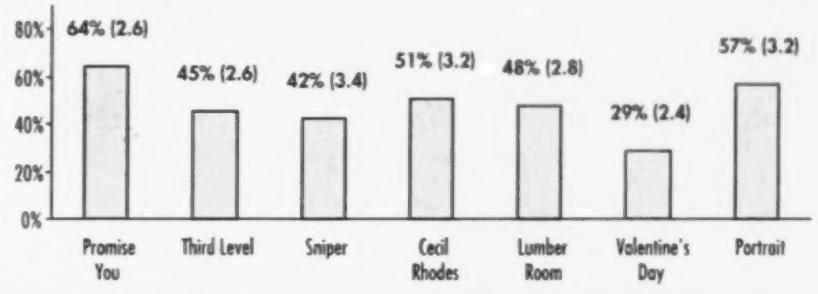


The standard errors of the estimated percentages appear in parentheses. It can be said with 95 percent certainty that for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP) 1994 Reading Assessment

For short constructed-response questions, the data presented are percentages of students who gave Acceptable responses. For extended constructed-response questions, the data presented are percentages of students who gave Essential or better responses.

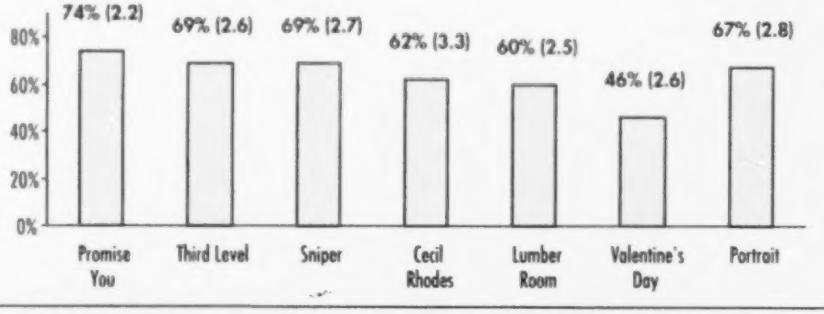
Figure D.2	**Question #3: *What do you find most meaningful about this story? Explain why.* Grade 12**	THE NATION'S REPORT CARD



The standard errors of the estimated percentages appear in parentheses. It can be said with 95 percent certainty that for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP) 1994 Reading Assessment

Figure D.3	**Questions #4: *What feeling was created by the language, setting, and major events in this story? Give specific examples.* Grade 12**	THE NATION'S REPORT CARD



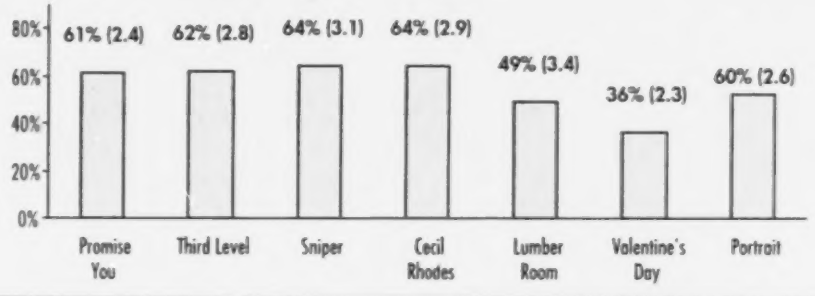
The standard errors of the estimated percentages appear in parentheses. It can be said with 95 percent certainty that for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP) 1994 Reading Assessment

---

For short constructed-response questions, the data presented are percentages of students who gave Acceptable responses. For extended constructed-response questions, the data presented are percentages of students who gave Essential or better responses.

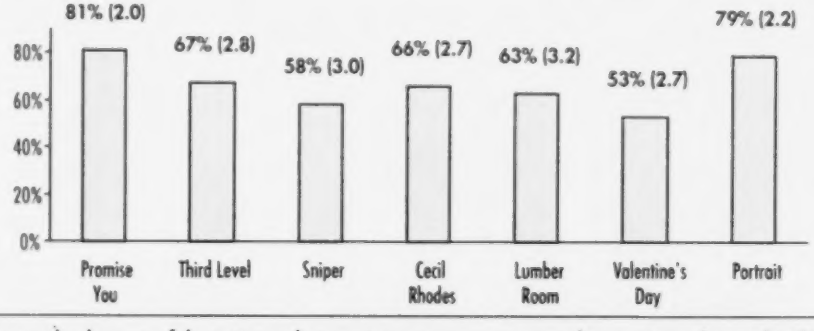
Figure D.4	**Question #5: *Could the events have taken place in another setting? Tell why or why not.* Grade 12**	THE NATION'S REPORT CARD



The standard errors of the estimated percentages appear in parentheses. It can be said with 95 percent certainty that for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP) 1994 Reading Assessment

Figure D.5	**Question #6: *Choose a conflict in this story and explain what the conflict is about.* Grade 12**	THE NATION'S REPORT CARD



The standard errors of the estimated percentages appear in parentheses. It can be said with 95 percent certainty that for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample.

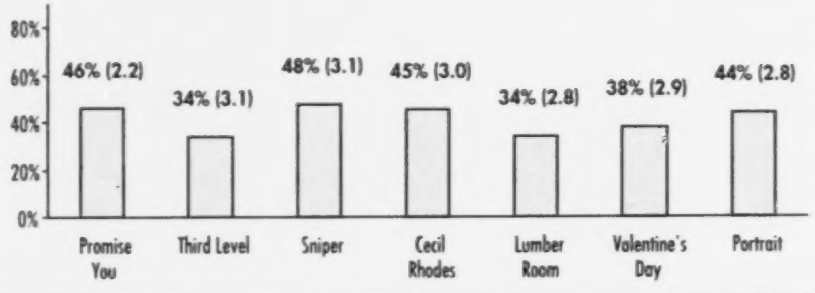
SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP) 1994 Reading Assessment

---

For short constructed-response questions, the data presented are percentages of students who gave Acceptable responses. For extended constructed-response questions, the data presented are percentages of students who gave Essential or better responses.

BEST COPY AVAILABLE

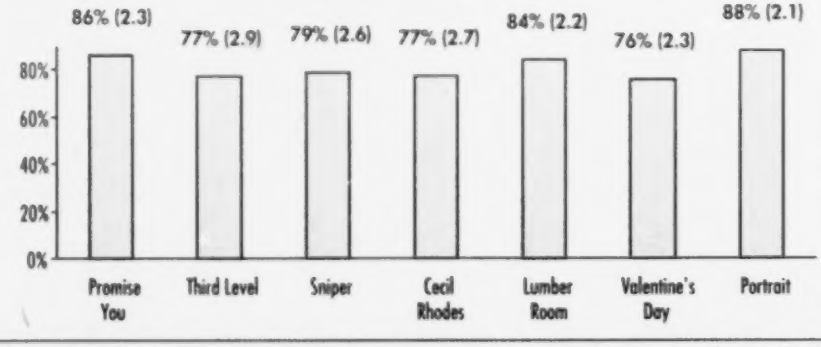
Figure D.6	**Question #7: *Discuss the most outstanding qualities of one of the main characters in this story. Support your answer with examples from this story* Grade 12**	THE NATION'S REPORT CARD



The standard errors of the estimated percentages appear in parentheses. It can be said with 95 percent certainty that for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP) 1994 Reading Assessment

Figure D.7	**Question #8: *Describe what happens in this story that causes or motivates one of the main characters to act the way he or she does.* Grade 12**	THE NATION'S REPORT CARD



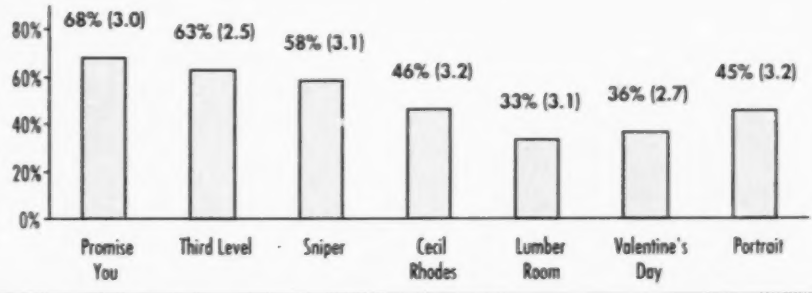
The standard errors of the estimated percentages appear in parentheses. It can be said with 95 percent certainty that for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP) 1994 Reading Assessment

---

For short constructed-response questions, the data presented are percentages of students who gave Acceptable responses. For extended constructed-response questions, the data presented are percentages of students who gave Essential or better responses.

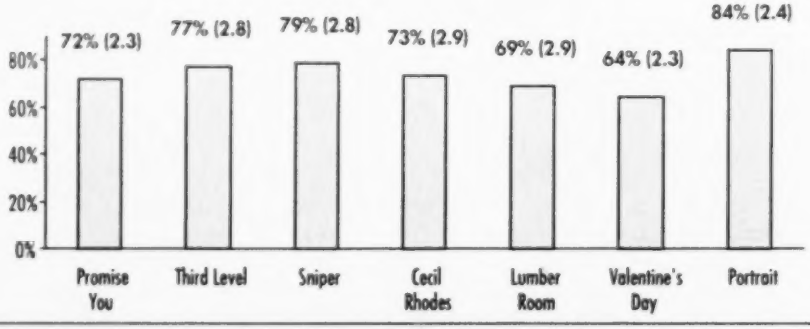
Figure D.8 **Question #9: *Choose an important event in the story and explain why it is crucial to this story's ending.* Grade 12**



The standard errors of the estimated percentages appear in parentheses. It can be said with 95 percent certainty that for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP) 1994 Reading Assessment

Figure D.9 **Question #10: Given what happened in this story, did you expect it to end the way it did? Give examples to support your answers. Grade 12**



The standard errors of the estimated percentages appear in parentheses. It can be said with 95 percent certainty that for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample.

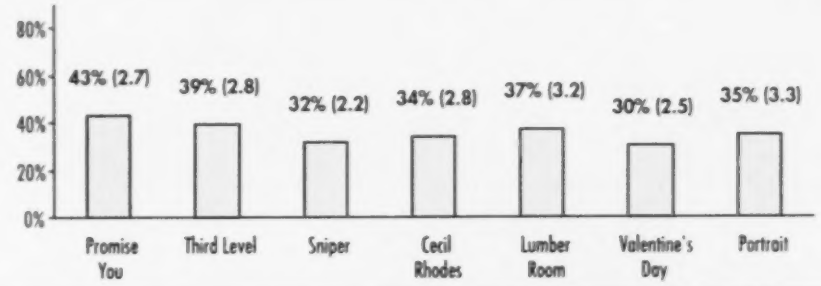
SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP) 1994 Reading Assessment

---

For short constructed-response questions, the data presented are percentages of students who gave Acceptable responses. For extended constructed-response questions, the data presented are percentages of students who gave Essential or better responses.

Figure D.10 **Question #11: *Tell how something in this story relates to something that happened to you.* Grade 12**

THE NATION'S REPORT CARD

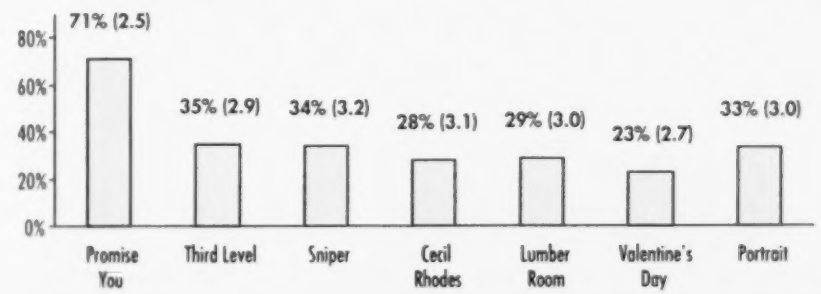


The standard errors of the estimated percentages appear in parentheses. It can be said with 95 percent certainty that for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP) 1994 Reading Assessment

Figure D.11 **Question #12: *What additional information would have helped you to understand this story?* Grade 12**

THE NATION'S REPORT CARD



The standard errors of the estimated percentages appear in parentheses. It can be said with 95 percent certainty that for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP) 1994 Reading Assessment

For short constructed-response questions, the data presented are percentages of students who gave Acceptable responses. For extended constructed-response questions, the data presented are percentages of students who gave Essential or better responses.

Acknowledgments

The work presented herein represents the efforts of the hundreds of individuals who are necessary to implement a complex special study of this size and scope. From the considerable expertise, energy, and dedication required to develop and conduct *the NAEP Reader* special study to that necessary to analyze and report it, many persons have made important and substantial contributions. Most importantly, NAEP is grateful to students and school staff who made the special study possible.

The NAEP Reader special study was truly a collaborative effort among staff from the National Center for Education Statistics (NCES), the National Assessment Governing Board (NAGB), Educational Testing Service (ETS), Westat, and National Computer Systems (NCS).

The special study was funded through NCES, in the Office of Educational Research and Improvement of the U.S. Department of Education. Emerson Elliot, past Commissioner, provided consistent support and guidance. The staff — particularly Gary Phillips, Peggy Carr, Susan Ahmed, and Sheida White — worked closely and collegially with ETS, Westat, and NCS staff and played a crucial role in all aspects of the program.

The members of the National Assessment Governing Board (NAGB) and the NAGB staff provided advice and guidance throughout. Their contractor for NAEP's reading consensus project, the Council of Chief State School Officers (CCSSO), worked diligently under tight time constraints to create the forward-looking framework underlying the assessment and the special studies.

NAEP owes a debt of gratitude to the consultants who provided their expertise and worked so conscientiously on developing *The NAEP Reader* instrument.

The NAEP project at ETS resides in the Center for Assessment of Educational Progress (CAEP) managed by Paul Williams. Under the NAEP contract to ETS, Steve Lazer managed test development activities. Jay Campbell worked with the Instrument Development Committee to develop the special study. Jules Goodison managed the operational aspects together with John Olson, and sampling and data collection activities were carried out by Westat under the direction of Renee Slobasky, Nancy Caldwell, and Keith Rust. Distribution, scoring and processing activities were conducted by NCS, under the supervision of John O'Neill, Judy Moyer, Diane Smrdel, and Brad Thayer.

Statistical and psychometric procedures were led by Nancy Allen. Steve Isham, with the assistance of Inge Novatkoski, performed *The NAEP Reader* data analysis. Jay Campbell and Patricia Donahue wrote the report. The design and production of the report was overseen by Carol Errickson, and the considerable production efforts were completed by Sharon Davis-Johnson and Barbette Tardugno. Word processing assistance was also provided by Terry Schoeps. The production of the World Wide Web version of this report was led by Pat O'Reilly with assistance from Laura Delate Miller and Debbie Kline.

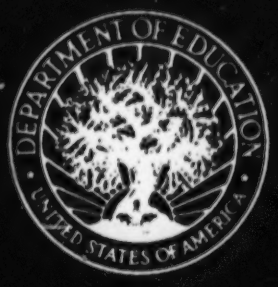
Many thanks are due to the numerous reviewers, both internal and external to NCES and ETS. The comments and critical feedback of the following reviewers are reflected in this report: Sue Ahmed, Nancy Allen, Marilyn Binkley, Peggy Carr, John Donoghue, John Guthrie, Debbie Kline, Andrew Kolstad, Steve Lazer, John Mazzeo, Michael Ross, Anne Sweet, and Sheida White.

United States
Department of Education
Washington, DC 20208–5653

Official Business
Penalty for Private Use, $300

Postage and Fees Paid
U.S. Department of Education
Permit No. G–17

Standard Mail (A)



NCES 97-491

106

BEST COPY AVAILABLE